DOCUMENT RESUME

ED 121 185 HE 007 623

AUTHOR	Nunn, Richard; Lain, Lindy
TITLE	Classification of Medical Education Institutions.
INSTITUTION	Association of American Medical Colleges, Washington, D. C.
SPONS AGENCY	Health Resources Administration (DHEW/PHS), Bethesda, Md. Bureau of Health Manpower.
PUB DATE	Dec 75
NOTE	87p.; Prepared by Division of Operational Studies
AVAILABLE FROM	Association of American Medical Colleges, Division of Operational Studies, One Dupont Circle, N.W., Washington, D.C. 20036
EDRS PRICE	MF-$0.83 HC-$4.67 Plus Postage
DESCRIPTORS	*Cluster Grouping; Factor Analysis; Grouping Procedures; *Higher Education; Literature Reviews; Medical Education; *Medical Schools; Research Methodology; *School Surveys; *Statistical Data
IDENTIFIERS	*Institutional Profile System

ABSTRACT

Empirical techniques are developed that may be used in conjunction with data stored in the Institutional Profile System to enhance present capabilities of assessing group structure in medical schools. Relevant literature is reviewed, and the institutionally descriptive data available for analysis and their manipulation into researchable formats are described. In order to relate the present data and methods to previous studies, variables similar to those used by the RAND Corporation were chosen. The data are factor analyzed, and the factor scores then used in two empirical cluster analysis procedures, and the results compared to those generated by RAND. Lack of replication is noted. The 10 clusters of medical institutions found by RAND were not found in this study. A variety of factors may have contributed to this conclusion, including a substantial number of both methodological and data differences between the studies. It is concluded that, at least for the present, categorization of medical schools by procedures accounting for multiple measures simultaneously does not yield clear and unambiguous results. This finding indicates that the picture presented by data institutionally descriptive of the schools is a highly complex one, not easily structured into a reasonably small number of groups of institutions. (LBH)

CLASSIFICATION OF MEDICAL EDUCATION INSTITUTIONS

PREPARED BY:

RICHARD NUNN
STAFF ASSOCIATE

AND

LINDY LAIN
RESEARCH ASSISTANT

ASSOCIATION OF AMERICAN MEDICAL COLLEGES
DIVISION OF OPERATIONAL STUDIES
ONE DUPONT CIRCLE, N.W.
WASHINGTON, D.C. 20036

DECEMBER 1975

CLASSIFICATION OF MEDICAL EDUCATION INSTITUTIONS

The work upon which this publication is based was supported in part by the Bureau of Health Manpower, Department of Health, Education and Welfare pursuant to contract number 231-75-0007. However, any conclusions and/or recommendations expressed herein do not necessarily represent the views of the supporting agency.

Chapter V - DISCUSSION AND CONCLUSION
by Douglas J. McRae, Ph.D.
Senior Staff Associate
Division of Operational Studies

Graphics by Debra Pitzer
Division of Operational Studies

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

This report, addressing Article I, Section 5, Part b(1) of the "Series of Analytical Studies on Medical Education and Academic Health Centers" contract between the Bureau of Health Manpower of DHEW and the Association of American Medical Colleges, describes the development and application of several cluster analysis techniques to data descriptive of U.S. medical schools for the purpose of classifying the schools into several categories or groups. The tasks set forth in the contract are as follows:

(a) A general classification methodology shall be developed to identify, ..., parameters which are manifest in available data and which reflect commonalities or dissimilarities across institutions.

(b) The methodological approach shall focus upon developing analytic clustering methods useful for classifying institutions on the basis of any type of empirical data. Such a scheme shall provide a series of classifications corresponding to the particular subsets of data used. Data subsets so visualized will include... faculty mix, student mix, and output characteristics.

(c) The classification structures will then be cross-validated against any other quantifiable data representing congruent information, other published research in the field, and verbal reactions by the medical education community, if available.

(d) Submit developed methodology in the form of a report.

This report is organized as follows. A review of relevant literature and an overview to the present study is given in Chapter I. In Chapter II, a description of the empirical cluster analysis techniques chosen for development and application is presented. A description of the institutionally descriptive data available for analysis and the manipulation of these data into researchable formats is given in Chapter III. The analysis of the data by empirical cluster analysis methods is presented in Chapter IV. Finally, Chapter V presents conclusions that may be drawn from the study and suggests steps for further analysis.

A. Review of the Literature

The need to classify medical institutions into a reasonably small number of groups is frequently voiced by those having to deal with U.S. medical schools in the process of policy development. There are currently 117 institutions in the U.S. at various stages of accreditation as medical schools. These 117 institutions present a diverse picture when viewed on institutionally descriptive measures, such as number and type of students, number and type of faculty, size and pattern of expenditures, curricula, and facilities.

In the absence of better schemes, schools are frequently classified on the basis of one or two selected measures. Classification by region and by type of ownership (public/private) are recurrent measures but provide limited insight into the real complexities of medical schools. A private school in the Western Region, for example, may be quite similar along many measurable dimensions to a public school in the Eastern region. Simple classificatory schemes tend to ignore such similarities. The present study is an attempt to develop classificatory methods capable of analyzing multiple measures simultaneously and subsequently grouping schools on the basis of similarities represented by those measures.

There have been several efforts to derive classificatory schemes for U.S. medical schools on institutionally descriptive measures. In particular, three recent studies are worthy of review.

Rodgers and Elton (1974), essentially replicating another study by Richards (1967), factor analyzed 14 variables descriptive of U.S. medical schools and then, based on the resulting factors, compared medical schools to one another through a technique known as "spatial configuration." Rogers identified the two factors "affluence" and "size", also found by Richards, but noted an additional factor labeled "graduate emphasis." To summarize the various scores attained by each medical school relative to these three factors, a plot of the "spatial configuration" was provided to illustrate the proximity (similarity/dissimilarity) of medical schools to one another as represented in two dimensional space.

The objective of the Otis study (1975) was to produce a general typology of U.S. medical schools for subsequent application in an analysis of "rates of production of differing types of physicians." Otis chose variables from several public sources and cluster analyzed related groups of them into five dimensions: size, eminence, clerkship versus basic science requirements, elective emphasis and services versus science funding. The individual scores for each medical school on these five dimensions were cluster analyzed by the BC-TRY object clustering routines (Tryon, 1970), producing ten medical school "types."

The RAND Corporation in 1972 conducted an extensive study of ten medical schools in the U.S. for purposes of a broader analysis of health manpower issues. In order to ensure that the ten schools selected were broadly representative of the entire population, multivariate cluster analysis was first applied to six factors (linear combination of variables) to form ten groups. From each of these ten groups, a single medical school was then selected. (Keeler, et. al., 1972.) This study used a methodological approach similar to that of the present study. As an application of the capability developed in this project, a classification similar to the RAND study was performed (Chapter IV). The RAND study is described in greater detail in Chapter IV.

B. Overview to the Current Study

The AAMC currently maintains in the Institutional Profile System a data base comprising over five thousand variables describing 117 U.S. medical schools. Coupled with this data source is a "user oriented" computer software package which offers a wide range of statistical and descriptive summary devices. This on-line system, which may be accessed through remote terminal sites, is intended to provide a facility for the exchange of information between members of the academic health community. It also provides a rich source of data for applied studies.

The general goal of this study is to develop empirical techniques that may be used in conjunction with data stored in the Institutional Profile System to enhance present capabilities of assessing group structure among medical schools. Primarily, the intention is to provide a means to group schools with similar profiles on a large number of measures. It should be pointed out, however, that there is no <u>one</u> cluster solution which will adequately characterize medical schools for all purposes; different solutions will be defined by different needs. One of the immediate objectives of this study is, therefore, to develop a methodology that may be used to augment the inter-institutional comparative methods currently available to users of the Institutional Profile System.

A detailed description of the type of methods available for such work and the methods chosen for implementation in this study is given in Chapter II. Empirical cluster analysis methods fall into a general category of statistically based techniques that may be labeled as applied multivariate descriptive analysis procedures. Other procedures falling into this category are factor analysis and multidimensional scaling. These procedures rarely yield exact, unequivocal results similar, for example, to probability statements that come from hypotheses testing statistical procedures. Rather, these techniques are better viewed as procedures that reduce highly complex multivariate data into simpler, perhaps, more revealing, formats.

On the substantive side of the study, variables on hand in the Institutional Profile System have been chosen and analyzed by empirical cluster analysis methods. An extensive set of variables (about 350) has been extracted from the IPS and prepared for research purposes. This set of data forms the basis not only for the present study but also for a large scale factor analytic descriptive study (Sherman, 1975) and a study of the effects of changes in class size (Sedlacek, 1975). This data extraction and variable manipulation represents the first large scale use of the Institutional Profile System for research in which applied multivariate methods have been used. The construction of this researchable data base is described in Chapter III.

The application of empirical cluster analysis methods to a subset of variables from the researchable set is described in Chapter IV. In order to relate the present data and methods to previous studies, variables similar to those used by the RAND Corporation have been chosen. Having factor analyzed the data, the factor scores are then employed in two empirical cluster analysis procedures, and the results compared to those generated by RAND. Differences between the RAND results and results of the present study are discussed in Chapter IV.

Chapter V presents conclusions concerning both the data and the methods as well as suggestions for further work. The cluster analysis procedures developed by this study are now available to users of the Institutional Profile System for application to selected subsets of variables.

Chapter II

METHODS

The term cluster analysis refers to a large body of methodological procedures designed to locate distinct groups of objects in which objects belonging to a group are in some way similar to each other but dissimilar to objects in other groups. The procedures available for such purposes range from highly subjective, judgement oriented methods to highly objective, statistically based methods. The cluster analysis procedures used for the present study come from the objective, statistically based end of this continuum, although some subjective judgements play a role in the results obtained.

This chapter describes the two approaches to cluster analysis used in the present study. It is important to understand that cluster analysis techniques are widely diverse and serve varied objectives. The benefits of one technique over another is realized only in light of the nature of the data in question and the purpose to which the results are to be put. These considerations in turn contribute to the operational meaning of the term "cluster". Given this background, the two approaches described in this chapter require somewhat different data attributes, and the results are interpreted accordingly.

Strictly interpreted, a key assumption of statistically based clustering procedures is that all objects must be placed into one and only one cluster. These procedures partition the entire set of objects (medical schools) into mutually exclusive and exhaustive subsets. This partitioning may take place whether the data sets are completely random or highly structured (i.e., whether or not there really <u>are</u> natural groupings). Outliers (unique objects similar to no other object) are either included in a cluster with other objects or constitute clusters by themselves. This conceptual constraint should be kept in mind when interpreting the results of any statistically based clustering procedure.

The two approaches to cluster analysis used in this study are in one case "hierarchical" and in the other "non-hierarchical". Each approach is described first, in general terms that include an illustration of typical results and then, in specific terms that detail the methodology chosen for this study.

A. Hierarchical Clustering Schemes

Hierarchical cluster analysis schemes generally construct groups of objects through a progression of stepwise merges. Initially, each object is considered a cluster in and of itself. A determination is then made as to which two clusters are most similar, whereupon these two clusters are merged. The process is then repeated until no further merge is possible.

This process starts with n objects or clusters, yields n-1 clusters after the first merge, n-2 clusters after the second merge, etc., until only one cluster (containing all n objects) remains. Hierarchical clustering schemes falling into this general framework have been labeled "agglomerative" hierarchical cluster analysis techniques*.

A feature of hierarchical, as opposed to non-hierarchical methods, is that once objects are grouped together they may not be separated later in the process. This feature offers both an advantage and a disadvantage. The early decisions greatly reduce the number of possible merges or changes that may take place later, thus allowing greater efficiency in the procedure. However, it precludes adjustment or reversal of unfortunate merges which have taken place earlier in the process.

Specific hierarchical clustering techniques differ from each other primarily in the criterion used to determine the basis for admittance of objects into clusters. These hierarchical procedures are described in this context in sections one and two below. In either case, an index depicting the status of the merge process, as plotted against the existing number of clusters, will often be helpful in determining an optimal solution found somewhere between the two extremes of n clusters

---

* There are a number of hierarchical techniques which work in a similar but reverse manner. They begin with a single cluster, containing all objects, and proceed to successively segment clusters into smaller and smaller groups. Such techniques are called "divisive" hierarchical cluster analysis procedures. They are not used in the study described in this paper.

(one object per cluster) and one cluster containing all n objects.

1. <u>Illustration</u>

To illustrate hierarchical clustering, consider the agglomerative procedure called the diameter method by Johnson (1967). This method fits into a general class of methods known as complete linkage. In the diameter method, the basic data analyzed is an n x n matrix of euclidian distances, where n is the number of objects to be clustered. At the first stage of clustering, the two objects with the smallest distance separating them are grouped together. At the next and all succeeding stages, an object is added to a cluster (note that a cluster may consist of only one object) only if the distance between it, the candidate object, and all objects within a cluster is less than its distance to all objects not in that cluster.

For example, four objects have the following distance matrix:

	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>
A	0	1.0	2.0	1.5
B		0	0.5	2.5
C			0	3.0
D				0

This matrix indicates that the distance between object A and object B is 1.0 units, that the distance between object B and object D is 2.5 units, etc.

At the first stage in Johnson's diameter method, objects B and C are grouped together because they are separated by the smallest distance in the matrix (i.e., they are most similar). At the second stage, the distance between object A and object D (1.5 units) is smaller than the distance between A and C (2.0 units); therefore, neither A nor D may be added to the B-C cluster and are grouped together to form the second cluster. At the final stage, the two clusters (B-C and A-D) are grouped together to form one cluster containing all four objects.

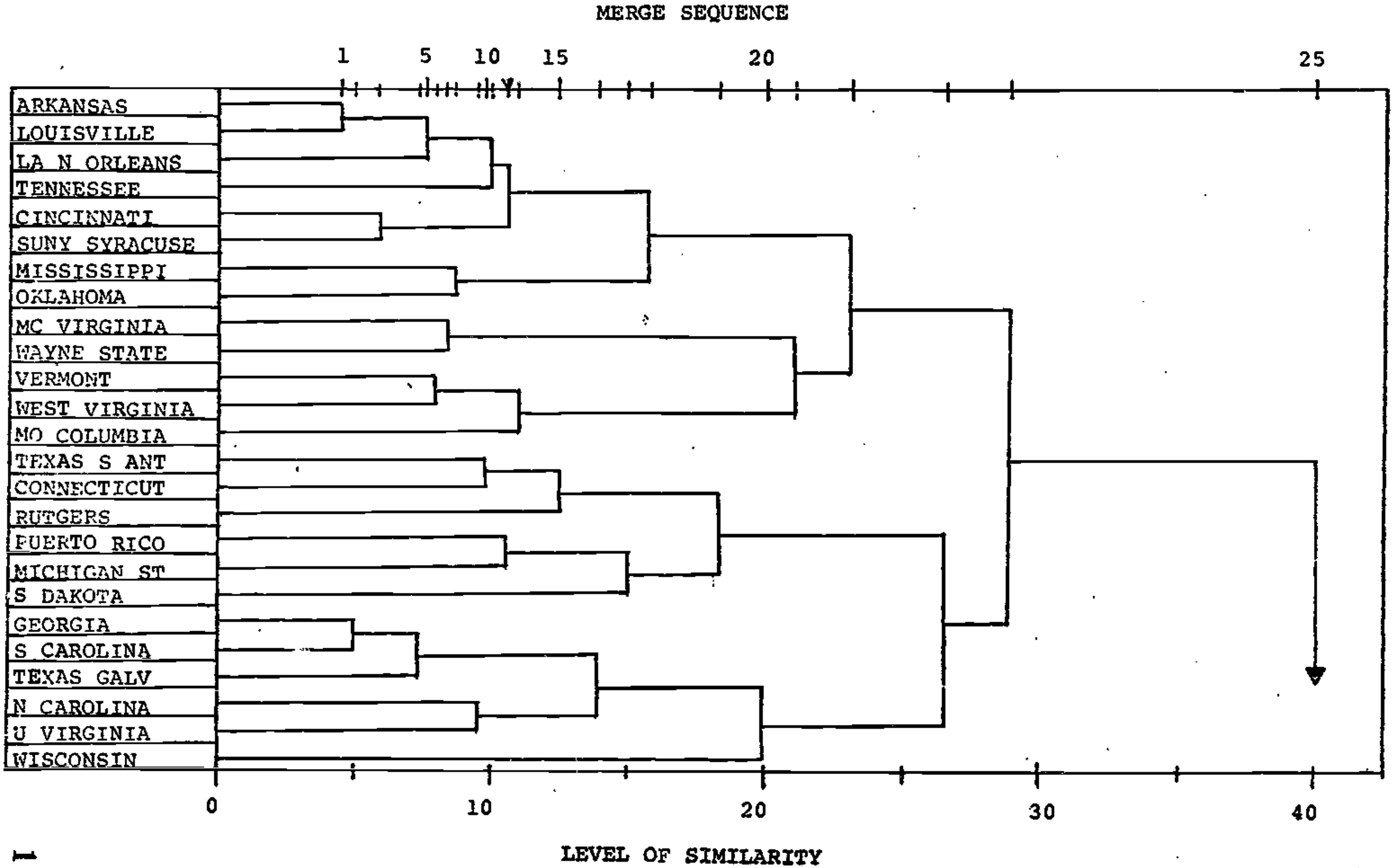
The merge criterion suggested by Johnson for this method is quite stringent and as a result produces clusters that are highly homogeneous. Although this characteristic may be beneficial under some circumstances, frequently, complete linkage methods are excessively constraining and fragmentary in their formulation of clusters. As Bailey summarizes: "complete linkage methods...dilate space. This means that the existing clusters move away from unclustered individuals as the clusters grow so that such individuals are more likely to form nuclei of new clusters than to add to pre-existing ones."

Despite this criticism, it is useful to examine the results produced by Johnson's diameter method because the method is simple and easy to interpret. To illustrate both the application of Johnson's diameter method and the typical graphic summary of results that normally accompanies hierarchical clustering, a dendrogram (tree diagram) is presented in Figure 2.1. Using data on each of 99 medical institutions for each of 6 factor scores (see Chapter IV for description of the factor scores), a 99 x 99 matrix of distances has been generated and submitted to analysis using the Johnson diameter method. Results shown in the figure are for 24 of the 99 institutions. The school names are listed on the left side of the page, and the critical distance for each merge is shown across the bottom of the figure. The sequence in which merges take place is recorded on the top of the figure.

A dendrogram may be interpreted by observing the development of linkages shown by series of interconnecting lines. Before a merge between any two schools takes place (at any point before merge sequence "1" is encountered), each school has a single line projecting horizontally to the right. At this stage there exists as many clusters as there are medical schools (25). The first merge, depicted by a link or vertical connecting line, occurs between Arkansas and Louisville. The distance separating these two schools can be determined by locating the corresponding position on the scale, "Level of

FIGURE 2.1

Dendrogram: Illustration of Hierarchical Cluster Formulation Based on Johnson's Diameter Method



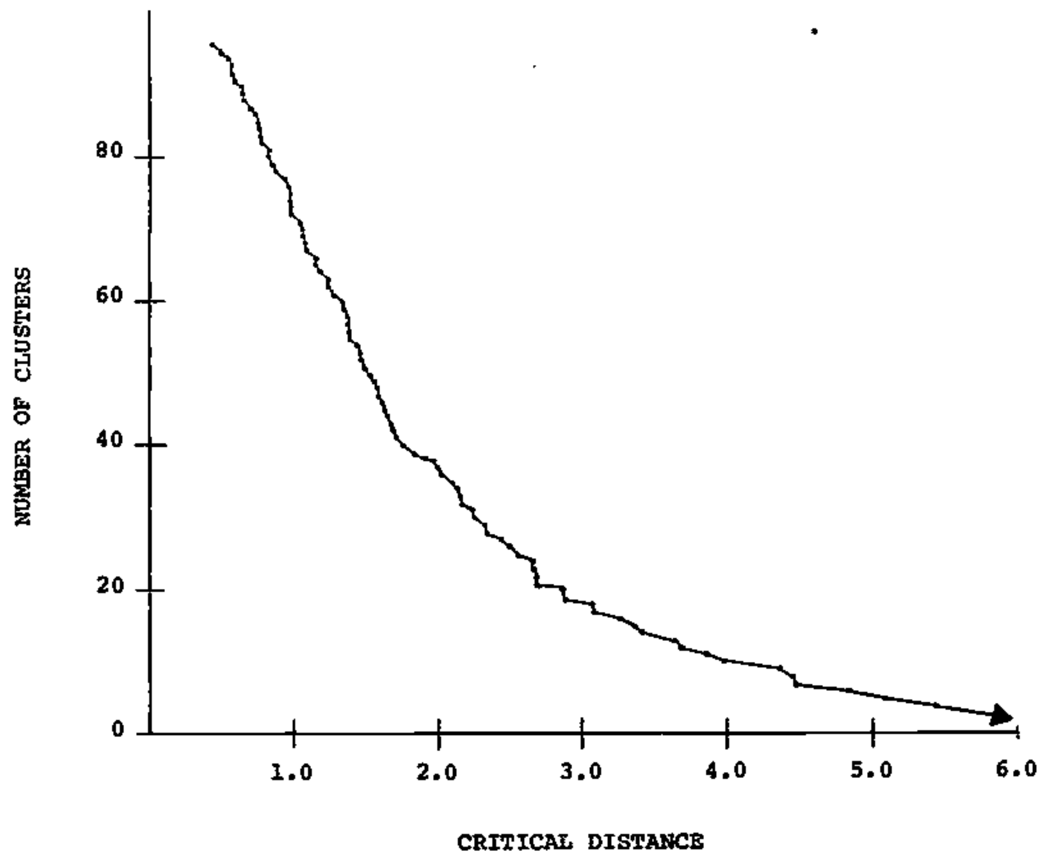
-14-

Similarity." In this case, Arkansas and Louisville are henceforth treated as a unit and, as such, are no longer denoted by individual horizontal lines but with a single line reflecting their joint status as members of the same cluster. The same steps may be applied throughout the dendrogram through n-1 merges, until all medical schools fall into one cluster. For purposes of illustration, a twenty-fifth merge has been included on the merge sequence scale. At this point in the clustering process, all 25 medical schools listed in the present dendrogram merge as a single group with another such group to form one cluster.

Since hierarchical techniques potentially present n-1 cluster solutions, some guidance is needed in selecting an "optimal" solution, i.e. to answer the question: "How many clusters are there?" This determination will often be more apparent when plotting the number of clusters existing at any given stage against the critical distance (or whatever criterion is used in the merging process). Such a plot, for the full 99 school analysis using the Johnson "diameter" method, is given in Figure 2.2.

This plot allows one to weigh the benefits of condensing clusters against the sacrifices to group cohension, expressed as critical distances, needed to facilitate a merge. The reduction in the number of clusters from 99 to 40, for instance, incurs only a slight relaxation in the critical distance. At the other extreme, merging into progressively fewer clusters entails

FIGURE 2.2


NUMBER OF CLUSTERS
80
60
40
20
0
1.0
2.0
3.0
4.0
5.0
6.0
CRITICAL DISTANCE

extending the critical distance disproportionately. Depending upon one's particular objective, an optimal solution is most likely found between 10 and 20 clusters.

2. Ward's Objective Function Technique

The hierarchical technique chosen for the present work is known as Ward's Objective Function method. More general than Johnson's diameter method, this approach "conserves" rather than "dilates" space. Rather than considering individual similarity measures between objects in the merging decision process, Ward's technique uses a general function based upon within-groups and between-groups "sum of squares". The general idea is to merge objects (or clusters) that produce the least increase to the within-groups or "error" sum of squares.

More specifically, one may calculate the sum of within-groups squared deviations as follows:

$$SS_W = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

where

SS_W = within-groups sum of squared deviations

X_{ij} = value for the jth object in the ith cluster

$\bar{X}_i$ = mean value for the ith group

n_i = number of objects in ith group

g = number of groups.

The within-group sum of squared deviations is essentially a measure of the collective compactness of the solution.

Cluster solutions with groups having members with highly similar profiles will yield low values for the within-groups sum of squared deviations.

At each stage of the process, the Ward objective function method merges those two objects which produce the least increase in SS_W. Stated another way, this method attempts to minimize within-cluster differences while maximizing between-cluster differences.

One feature of the Ward method is that the centroid for each cluster changes after each merge takes place. Thus, after a merge, the values for the X's change (if, for no other reason than the number of clusters is constantly decreasing). This dynamic property may be viewed as both a benefit and a drawback to the method. On one hand, this property permits a more realistic approximation of the current composition of members within clusters. On the other hand, it tends to allow centroids to migrate towards outlying objects that are forced into clusters by virtue of the mutually exclusive and exhaustive nature of the clustering process. The migrating centroid effect may cause objects to be included in existing clusters rather than to be formed into new ones.

The Ward method is used extensively in the present study. Although the results are similar on the surface to those from the Johnson method illustrated in Figure 2.1, it is important to keep in mind that the Johnson and Ward methodologies are quite distinct. The Ward method was chosen because of its compatibility with the underlying assumptions and objectives of this study, and, moreover, its compatibility with a non-hierarchical technique also used in this study.

B. Non-Hierarchical Clustering Schemes

Unlike hierarchical clustering, non-hierarchical clustering does not develop clusters through a progression of step-wise merges. Instead, the user typically indicates the number of clusters to be formed. The non-hierarchical technique attempts, then, to place all objects into the specified number of clusters in order to optimize a given criterion. Most frequently, the criterion is the SS_W described above, although other criteria have been suggested (Friedman and Rubin, 1967).

After the set of objects to be clustered is initially broken into a desired number (specified in advance) of partitions and objects are assigned to groups in either a systematic or arbitrary fashion, non-hierarchical procedures then proceed to reassign objects to that cluster which most closely approximates the objective criterion. The procedures for initially partitioning and then reassigning objects in order to optimize a criterion provide for the variety of specific non-hierarchical methods that have been suggested in the literature (Ball and Hall, MacQueen, Forgy, Jancy, McRae, Friedman and Rubin).

1. Illustration

To simplify the steps involved in non-hierarchical clustering, an illustration is provided in Figure 2.3. The data set used in this example, set forth in step 1, comprises ten objects, lettered A through J, and two variables, X and Y. As pointed out above, non-hierarchical techniques often require an advance specification as to the number of clusters into which objects are to be sorted. For each cluster specified, the user will normally supply a seedpoint, which is merely a point in the measurement space around which clusters are expected to materialize. Since the present example is two-dimensional, a plot of the ten objects relative to the variables X and Y makes the task of specifying a suitable number of clusters and the location of their respective seedpoints considerably easier. The proximity of these ten objects represented in space suggests the presence of three clusters. They are, as outlined in 2, a group consisting of objects C, F, I and A, another with objects H, D, and E, and another with J, G, and B. The seedpoints, denoted by triangular marks, have been situated in such a way as to approximate centers of these clusters.

The last two steps constitute an effort to improve the original estimates of seedpoints and to adjust the member composition of each cluster accordingly. How these steps are accomplished is essentially defined by the non-hierarchical

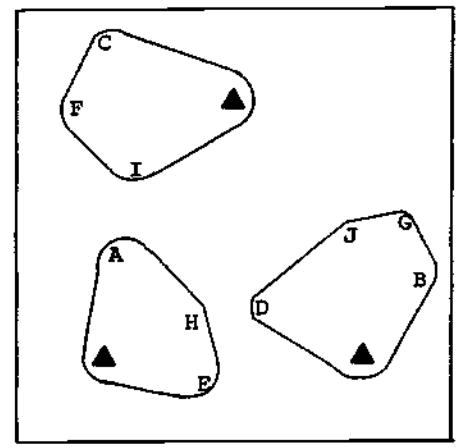
FIGURE 2.3

An Overview of Steps Undertaken in Nonhierarchical Cluster Analysis

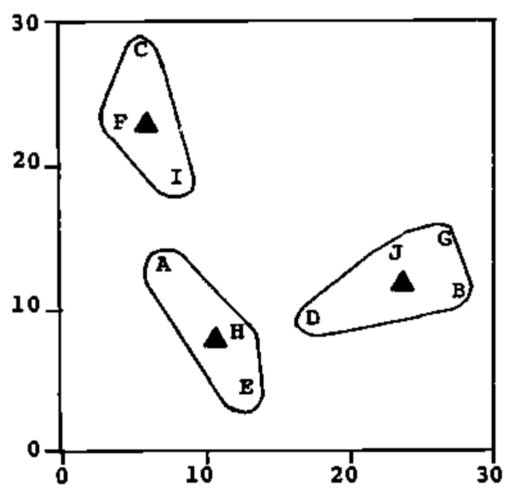
1) THE DATA SET

ID	VARIABLE X	VARIABLE Y
A	7	13
B	28	11
C	6	28
D	17	9
E	13	4
F	4	23
G	27	15
H	12	8
I	8	19
J	23	14

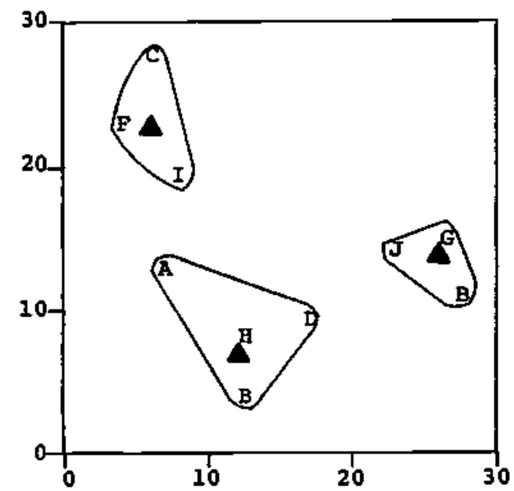
2) SELECT SEEDPOINTS TO SPAN THE DATA SET - ASSIGN ALL OBJECTS TO THE NEAREST SEEDPOINT



3) ESTABLISH CLUSTER CENTROIDS BY UPDATING SEEDPOINTS



4) REASSIGN OBJECTS TO NEAREST CENTROID AND UPDATE



algorithm itself. One approach may be first to recompute cluster centroids by establishing an average or central point for each group of objects as tentatively assigned in 2. All objects are then reassigned to that cluster having the nearest centroid. It is important to note that step 3 may result in significant alterations (especially without the benefit of advance knowledge of the data structure) in the original estimate of cluster centroids and membership. Object D, for instance, has changed its cluster affiliation between steps 3 and 4. Essentially, step 4 involves a repeat of step 3, again for the purpose of refinement. In this example, no further adjustments prove necessary beyond step 4 because no change occurs in cluster membership after the centroids have been updated.

2. Forgy's K-means Technique

One of the earliest non-hierarchical techniques proposed was the K-means approach by MacQueen. With this technique, the user specifies the number of clusters to be generated, for example, g. Then the first g objects in a data set are arbitrarily taken as representing the centroids for g clusters. The remaining objects are considered in sequence and assigned to the cluster whose centroid is least distant. After each assignment, the cluster centroid is recalculated to reflect the last entry. When all objects have been assigned to groups, cluster centroids remain fixed. Because the original seed-points have been updated at each entry of an object to a cluster, a final pass through the data is made in order to reassign objects that have become closer to other centroids. The final pass requires no further updating of the centroids.

As originally proposed by MacQueen, K-means is a two pass procedure. The first pass, just described, finds centroids; a second pass makes final assignments of objects to clusters. It is important to note that objects assigned to clusters on the first pass may be assigned to different clusters on the second pass. As indicated above, this is one property that distinguishes non-hierarchical techniques from hierarchical techniques.

It should also be noted that by assigning objects to clusters based upon smallest distances, the K-means technique is very similar to the Ward method which attempts to minimize the within-group sum of squares. MacQueen (1967) presents theoretical and empirical evidence of this similarity. Thus, the K-means procedure and Ward's objective function procedure share similar objectives but differ primarily in the arbitrary procedures designed to achieve these objectives.

Forgy (1967) suggests modification to MacQueen's basic K-means method in two substantial ways. First, he suggests that the process continue iterating as long as an objective function, such as the within-groups sum of squares, continues to decrease. Secondly, he suggests that centroids not be recalculated until the end of each iteration. These two changes result in a non-hierarchical process similar to the Ward method in its formal attempts to minimize an objective function. The Forgy modifications, however, overcome two of the major problems of the Ward procedure, the permanence of cluster membership inherent in the hierarchical approach and the difficulties associated with migrating means.

In all non-hierarchical procedures, the specification of initial cluster centroids (seed points) is of great importance. This specification may be done randomly as in MacQueen's method in which the <u>first</u> g objects are taken, or it may be based upon some substantive grounds. In particular, if the investigator has some notion of where concentrations of objects occur in the structure, he may wish to ensure that clusters be given every chance to grow in that area. Thus, one modification to the Forgy procedure that may be used, if sufficient knowledge of the data structure is available, is that of specifying values to serve as initial seedpoints.

One further observation, although both the Ward technique and Forgy technique attempt to minimize the within-group sum of squares, there is no guarantee that either technique will reach the absolute minimum. The only way to ensure the attainment of an absolute minimum sum of squares is through complete enumeration of the data set, but even aided by today's advanced computer technology, a complete enumeration is unrealistic for all but very small data sets.

The two clustering procedures chosen for the present study are used in successive stages. First, a hierarchical cluster analysis, Ward's objective function, is used. Second, results from Ward's method provide seed points for analysis of the same data by Forgy's non-hierarchical procedure. Computer programs for both of these procedures have been obtained and adopted for use on the AAMC's Institutional Profile System. The use of these procedures is illustrated in Chapter IV. The procedures are now available for application to subsets of data in the Institutional Profile database.

Chapter III

DATA

Beyond variations evident in the objectives defined by clustering methodologies, the properties assumed in any configuration of clusters is essentially a reflection of the data employed. This chapter focuses on the development of a data set suitable for use with cluster analysis. The two areas to be considered are, first, the availability and selection of variables from the Institutional Profile System and, second, the preparation of the data for analysis.

The Institutional Profile System (IPS) is a computerized information retrieval system with a large data storage capacity and software to perform various statistical and data summary functions. Currently, the data base includes data on U.S. medical schools for 16 years and contains over five thousand variables from 49 source questionnaires. The sources of interest for the present study are primarily the Liaison Committee on Medical Education Questionnaire: Parts I and II, 1973-74.

A. The Data Set: Availability and Selection

The principle objective in developing a data set was to assemble a comprehensive set of variables with a sufficiently broad, yet detailed, perspective of medical education to facilitate exploratory analyses. In addition to this study, two other exploratory studies under this contract also required utilization of such a data set. (Sherman, 1975; Sedlacek, 1975). Basically, then, the variables in the data set are intended to have a general rather than contextual descriptive value.

The contract specification related to this study, provides for the development of cluster analytic methods that may be used as an additional function within the IPS statistical package. Because, however, the cluster algorithms made available to the organization are not compatible with the program language used in IPS, all analyses conducted for this paper were done external to IPS. The first step, then, involved extraction of relevent data from IPS in order that these analyses might be performed. The studies conducted by Sherman and Sedlacek also required an external application of the data for use in the statistical programs available in SPSS *
(Nie, 1975).

Variable selection involved the identification of the most current and meaningful data available. Selection began with the most current IPS data for 117 medical institutions. Although the bulk of the data was available for academic year

---

* Statistical Package for the Social Sciences

1973-74, the most current financial data was for 1972-73.

In attempting to formulate a full spectrum of salient institutional descriptors, suggestions were elicited from AAMC staff representing a number of specific areas in academic medicine. Other potentially useful descriptors were noted in the studies by Richards (1966), Rodgers, Otis, and RAND and were organized into logical domains given in Table 3.1. To facilitate comparison, variables used by each study are identified by an X in the columns to the right. The emphasis of each study on particular variable domains, as shown in this table, varies considerably. The most obvious omissions occur in the faculty domain, with the exception of RAND, and in the curriculum domain for all but Otis' study.

The more current and extensive data available in IPS allow an expansion of the data set for the present study to 350 variables. The set contains approximately 220 variables taken directly from source documents and an additional 130 variables derived from the original 220, (mostly ratios and percents). The entire set is listed in Appendix A. A summary of the expanded domains is given in Table 3.2.

B. Data Preparation

The development of a data set, particularly one of this size, requires a number of preliminary tasks. These tasks include organizational considerations, such

TABLE 3.1

	OTIS	RICHARDS	RODGERS	ARD	
I. THE INSTITUTION					
A. ORGANIZATION, PHYSICAL FACILITIES, SETTING AND GENERAL CHARACTERISTICS					
1. library volumes per student	X				
2. Ratio of number of beds in teaching hospital to no. of medical students		X			
3. % of total beds in university hospital				X	
4. Private vs. public	X	X	X	X	
5. Age of Institution	X	X		X	
6. Growth rate		X			
7. Size of community in which located		X			
B. ADMISSIONS					
1. no. of applicants per place available	X				
2. % of male applicants	X				
3. average no. of applications per applicant	X				
4. accept transfer students	X				
5. % of out-of-state students	X	X	X	X	
6. Ratio of entering to applying students		X	X		
7. % of foreign students in entering class		X			
8. % of part-time and special students in student body		X			
9. % of entering students completing 4 years of college		X	X		
C. FINANCES					
1. Deci'e Federal Research funds	X				
2. Dollars from sponsored programs per student	X				
3. Deci'e total Federal sources of support	X				
4. Deci'e unrestricted endowment funds	X			X	
5. % of schools HEW contribution for science research	X				
6. % total expenditures for sponsored programs	X			X	
7. % of schools HEW contribution for all other non-science	X				
8. % of schools HEW contribution for science training	X				
9. % total expenditures for regular operating budget	X				
10. % of schools HEW contribution for non-science training	X				
11. % of schools HEW contribution from environmental health services	X				
12. % of schools HEW contribution from health services and mental health administration	X				
13. % of total federal obligations.			X		
14. % of schools HEW contributions from NIH	X				
15. % of schools HEW contribution for science training	X				
16. Private funding sources	X				
17. Tuition cost	X	X	X		
II. THE FACULTY:					
A. COMPOSITION (FACULTY MIX)					
1. No. of full-time faculty	X				
2. Ratio of part-time faculty to full-time faculty				X	
3. Ratio of volunteer faculty to full-time faculty				X	
B. SALARY					
C. FINANCES					
1. Research funds per faculty member	X				
2. % of faculty salary from Federal dollars				X	
3. Sponsored program expenditures per full-time faculty				X	
D. PROFESSIONAL EMPHASIS					
1. faculty per student ratio	X			X	
2. % teaching responsibility for clinical fellows	X				

	OTIS	RICHARDS	RODGERS	RAND	********
III. THE STUDENT:					
A. COMPOSITION (STUDENT MIX)					
1. Decide MCAT science score	X			X	
2. % of males in final year	X				
3. % of first year students of student body			X		
4. ratio of no. final year students to first year students			X		
5. no. of graduates			X		
6. total enrollment in post doctoral B.S. program			X		
7. interns in major teaching hospitals				X	
8. no. residents in major teaching hospitals				X	
9. ratio of interns and residents to medical students				X	
10. ratio of masters and doctorates in B.S. to medical students				X	
11. ratio of student equivalents to medical students				X	
12. % males in student body		X	X		
13. no. graduate degree candidates in B.S.		X			
14. no. post doctorate fellows in B.S. and C.S.		X			
15. ratio of interns to medical students		X			
16. ratio of residents to medical students		X			
B. STUDENT AID					
1. financial aid		X			
C. FINANCES					
1. Total expenditures per student	X			X	
2. Dollars training support per student	X				
3. Regular operating expenses over total students				X	
4. Expenses for books and supplies for first year students			X		
D. CURRICULUM AND PROGRAM					
1. No. of residency programs	X				
2. No. of types of residency programs	X				
3. No. of intern programs	X				
4. Weeks of required clerkship	X				
5. % of instruction devoted to B.S. requirements	X				
6. % of instruction devoted to clerkship requirements	X				
7. Year required clerkship introduced	X				
8. Total weeks of instruction	X				
9. No. of types of internship programs	X				
10. elective emphasis	X				
11. % elective time	X				
12. all elective final year	X				
E. ENROLLMENT					
1. Total size/enrollment	X	X	X	X	
2. Size of first year class	X				
3. Size of final year class	X				
4. No. students per administrative official	X				
5. Total students (affiliated)				X	
F. OUTPUT CHARACTERISTICS					
1. Specialty Board Certification rate	X				
2. Residency preference	X				
3. Completion rate/attrition		X			
4. Ratio of doctorates conferred to total enrollment			X		

Table 3.2

	No. Variables *
I. INSTITUTION	(22)
A. General Characteristics	14
B. Demographic	5
C. Library Facilities	3
II. FINANCES	(86)
A. Revenues	37
B. Expenditures	21
C. NIH Awards	6
D. Construction Costs	14
E. General	8
III. ACADEMIC PROGRAM	(39)
A. General	11
B. Curriculum	28
IV. FACULTY	(48)
A. Staff	32
B. Salary	16
V. STUDENT ADMISSIONS	(164)
A. Enrollment	69
B. Entering Qualifications	30
C. Student Aid	40
D. Expenses	6
E. Student Selection	14
F. Career Review	5

* Parentheses denote sub-totals per variable domain

as the formating, labeling, transformation, and storage of data. Additional steps entail the creation of new measures from existing variables and, finally, verification of the entire data set.

Because of the operational benefits provided by the statistical programming package, SPSS, the IPS interface function was used to remove data from the system prior to undertaking these preliminary steps. On the basis of the 220 variables extracted from ISP, an additional 130 derived measures were computed. Most of these measures represented the creation of percentages and ratios. The final preparatory step involved performing univariate frequency tabulations and summary statistics on the 350 variables. These computations provided the basic documentation needed for verifying the substance of the data. Additionally, an analysis of the incidence of missing data for the entire selection of variables was performed. The results indicated that while the overall incidence of missing data was negligible, it did occur in high concentrations among ten percent of the institutional population. The findings in these data summaries were treated separately according to the needs of this study and the Sherman and Sedlacek studies.

Chapter IV

APPLICATION

In this chapter, the Ward and Forgy clustering methods will be applied to select variables extracted from the database described in Chapter III. The variables have been specifically chosen to closely approximate those used by the RAND Corporation (Keeler, et al, 1972) in a study designed to classify medical schools. This chapter, then, is an attempt to verify the RAND study. Such an effort is expected to shed light on three sources of concern: (1) to test the adequacy of the methods developed by AAMC and by RAND, (2) to test the adequacy of the data analyzed by AAMC and by RAND, and (3) to detect possible changes in medical education over time as reflected by the measures analyzed.

A. The RAND Study

In 1972, the RAND Corporation was commissioned to conduct a broad study of the effects of federal programs on academic health centers. The project initially required a selection of ten medical schools that would be representative of all medical schools in the United States. To accomplish this task, RAND researchers selected these institutions by classifying medical schools into ten groups and choosing one school from each group for study.

The RAND study utilized classificatory methods similar to those presented in Chapter II. They selected 31 variables deemed broadly descriptive of medical education and obtained data for 94 medical schools. The first phase of their analysis involved a factor analysis * of the 31 variables which, in turn, yielded six common factors.

Factor scores were then computed for each of the six factors for each institution and submitted to non-hierarchical cluster analysis for which ten clusters were specified. The results of this analysis are presented in Table 4.1.

To summarize, RAND conducted factor and cluster analyses, first, to isolate underlying dimensions existing in their selection of variables and, second, to identify distinct groups on which to base a representative selection of medical schools.

B. Replication: Factor Analysis

Replication of the RAND Study involves two distinct steps. The first step is undertaken in Sherman's study (1975) in which 23 variables comparable to RAND's 31 are submitted to factor analysis. The second step, detailed in this chapter, is to cluster medical schools based on the six factors identified by Sherman.

A list of variables used by Sherman and RAND are provided in Table 4.2. By utilizing the same set of procedures as did

---

* Common factor analysis, followed by equimax rotation.

TABLE 4.1

CLUSTER 1 (13 MEMBERS)

Oregon
Ohio State
Colorado
Kentucky
LA, New Orleans
Tennessee
Minnesota
Med College of GA,A
Arkansas
Kansas
Texas, Southwestern
SUNY, Buffalo
Indiana

CLUSTER 2 (5 MEMBERS)

UC-Davis
Michigan State
LA, Shreveport
UC-Irvine
Mount Sinai

CLUSTER 3 (11 MEMBERS)

Med C of VA
Maryland
Med College of Wisconsin
Northwestern
Wayne State
SUNY, Downstate
Hahnemann
Thomas Jefferson
Illinois
Loma Linda
U of Michigan

CLUSTER 4 (10 MEMBERS)

Case Western
Columbia
U of Pennsylvania
NYU
UCLA
UCSF
Harvard
Yeshiva, Einstein
U of Washington
USC

CLUSTER 5 (13 MEMBERS)

Oklahoma
Puerto Rico
Vermont
SC
U of VA
Mississippi
UNC
Louisville
Missouri
Nebraska
West Virginia
Iowa
U of Wisconsin

CLUSTER 6 (3 MEMBERS)

Med College of Ohio
UC-San Diego
Arizona

CLUSTER 7 (12 MEMBERS)

Pittsburgh
Cincinnati
NJ Med School
Temple
SUNY, Upstate
Bowman Gray
Miami
Florida, Gainsville
Cornell
Texas, Galveston
Texas, San Antonio
Penn State

CLUSTER 8 (10 MEMBERS)

Yale
Washington, St. Louis
Emory
Johns Hopkins
Stanford
Duke
Vanderbilt
Rochester
Baylor
U of Chicago

CLUSTER 9 (13 MEMBERS)

Tulane
Georgetown
Med C of PA
Boston
Loyola, Chicago
Albany
Saint Louis
NY Med
Chicago Med
Tufts
Howard
George Washington
Creighton

CLUSTER 10 (4 MEMBERS)

Utah
Alabama
U of New Mexico
Meharry

TABLE 4.2

Variables Used in Replicated Factor Analyses

RAND (1972)

1. Medical Students
2. Interns in Major Teaching Hospitals
3. Residents in Major Teaching Hospitals
4. State or Private School Status
5. Unrestricted Endowment (decile)
6. MCAT Science Scores (decile)
7. Percent Faculty Salary from Fed & (decile)
8. State Medicaid Program
9. Percent NIH Research Applications Approved
10. Average Priority Score
11. Population SMSA/Total Medical Students SMSA
12. NIH Research and Training Grant $ (FY 1971)
13. Total Students
14. Percent of Medical Students from Home State
15. Special Project $/Total Students
16. Log (1972 - year organized)
17. Percent of Total Beds in University Hospital
18. Percent of Total Beds in VA Hospital
19. Part-time Faculty/Full-time Faculty
20. Volunteer Faculty/Full-time Faculty
21. Full-time Faculty/Total Students
22. Sponsored Program Expenditures/Full-time Faculty
23. Regular Operating Expenditures/Total Students
24. Total Expenditures/Total Students
25. Sponsored Program Expenditures/Total Expenditures
26. (Interns & Residents)/Medical Students
27. (Masters & Doc. in Basic Science)/Medical Students
28. Financial Distress $/Regular Operating Expenditures
29. $ Weighted Priority Score - Priority Score
30. $ Weighted Fraction Approved - Fraction Approved
31. Other Student Equivalents/Medical Students

AAMC'S REPLICATION (1975)

1. Medical Students (73-74)
2. Total Interns Instrcuted by MC Faculty (72-73)
3. Residents Instructed by MC Faculty (73-74)
4. Public or Private Control (73-74)
5. Tot MC Rev from Unrestricted Endowments (72-73)

6. MCAT Science Scores of 1st Year Medical Student (73-74)
7. Percent Sponsored Faculty Salary from Federal $ (72-73)
8. SMSA Population per Medical Student (73-74)
9. NIH Awards - Research Grants $ (73-74)
10. Total of all Students Instructed at MC (73-74)

11. Percent of Medical Students from Home State (73-74)
12. Special Project $ per MD Students (72-73)
13. Age. Log (1974 - year organized)
14. Part-time Faculty/Full-time Faculty
15. Volunteer Faculty/Full-time Faculty

16. Full-time Faculty/Total Students (73-74)
17. Sponsored Program Expenditures/Full-time Faculty
18. Regular Operating Costs per MD Student (72-73)
19. Total Expenditures/Total Students (73-74)
20. Sponsored Program Expenditures/Total Expenditures

21. (Interns & Residents)/Medical Students 73-74
22. (Masters & Doc. in Basic Science)/Med Students
23. Medical Student Equivalents/Medical Student (73-74)

RAND, that of common factor extraction and equimax rotation, Sherman finds essentially the same six factors based on 117 institutions as opposed to RAND's 94. They are, as presented in Table 4.3: (1) graduate medical education programs, (2) Federal research involvement, (3) undergraduate medical education programs, (4) reliance on non-full-time faculty, (5) public versus private control and (6) non-M.D. education programs.

A factor, which may be viewed as simply a synthetic variable, is a condensation of a group of variables into a single expression. Each school's position along the descriptive dimension represented by a "factor" (such as the second, "Federal research involvement") is given a "factor score" computed from the input variables using a formula derived by the factor analysis. There are six such formulae, one for each factor. Each school, then, has six factor scores that replace values for the 23 vairables. The computed similarity of two schools is a composite measure of the similarity of their six factor scores. This composite measure includes a set of numerical weights to reflect the subjective importance of each of the six dimensions in determining the school's similarity. The present study uses the same numerical weights assigned by RAND.

TABLE 4.3

Factor Pattern Matrix from Analysis of
RAND Study Variables (Using New AAMC Data)
By Method of Common Factors and Equimax Rotation

VARIABLES	RAND FACTOR LABELS	VARIABLE LABELS	FACTOR (VARIABLE GROUPS) 1	2	3	4	5	6
V6310	Graduate Medical Education Programs	TOT RESDNTS INSTR BY MD FAC 73-74	.86	.22	.28	-.03	.07	-.03
V6330		TOT INTERNS INSTR BY MD FAC 72-73	.80	.19	.21	-.02	.05	-.00
V6080		ENROL RATIO-INTERNS & RESDNTS TO MD STU	.64	-.10	-.29	.09	.11	.22
V6010		TOT STUDENTS...ALL...INSTRUCTED AT MC	.61	.38*	.60	-.17	-.12	.21
V3350	Federal Research Involvement	SPONS PROG EXPD PER FT FAC	.10	.86	.19	-.00	.11	.09
V3345		MC EXPD-REG OP COSTS PER MD STUDENT	.24	.61	-.19	.50	.09	.17
V2830		MC EXPD-PCT SPONS PROG EXPD OF TOT	.04	.67	.25	-.06	.34*	.12
V2940		NIH AWARDS RESRCH GRANTS $1000 73-74	.42*	.58	.06	.35*	.30	.36*
V7200		MEAN MCAT SCORE SCI-1ST YR MD STUDENTS	.38*	.41	.00	.09	.23	.30
V2820		PCT SPONS FAC SALARY FROM FED $ 72-73	-.14	-.32	-.12	.02	-.20	.26
V6020	U.G. Med. Educ. Programs	ENROL'L-TOT MD STUDENTS 73-74	.30	.22	.85	-.08	.02	-.11
V1045		AGE OF INSTITUTION	.08	.08	.74	-.07	.29	.04
V1140		SMSA POP PER MD STUDENT	.17	.06	-.33	.11	-.05	-.27
V5025	Reliance on Non-Full-Time Faculty	RATIO FT FAC TO TOTAL STUDENTS	-.11	-.02	-.29	.76	.05	.04
V2750		TOT MC EXPD PER TOTAL STUDENTS	-.19	.52	-.12	.81	.06	-.07
V5040		RATIO VOL FAC TO FT FAC	.05	.02	-.45*	-.41	-.16	-.37
V2740		SPECIAL PROJ $ PER MD STUDENT 72-73	-.20	.11	-.11	-.34	-.00	.01
V1030	Control: Public Vs. Private	CONTROL TYPE (1=PRIVATE, 0=PUBLIC)	.10	.05	.01	.07	.78	-.10
V6230		PCT MD STUDENT FROM HOME STATE	.11	-.20	-.24	.16	-.58	-.09
V2110		MC REV-TOT UNRESTR ENDOW & GIFTS	.14	.17	.06	.29	.56	.11
V6050	Non-M.D. Educ. Programs	ENROLL RATIO-MD STUDENT EQUIV TO MD STU	.08	.12	.16	.24	-.29	.64
V6140		ENROLL RATIO-MAS & DOC BAS SCI TO MD STU	.17	.16	.00	.07	.07	.52
V5030		RATIO PT FAC TO FT FAC	.03	.01	.03	-.30	-.20	-.43

C. Replication: Cluster Analysis

1. Analysis by Ward's Objective Function Method

Based on the six factors identified by Sherman, two independent cluster analyses are conducted. The first cluster application, which is to be discussed in the immediate section, is the Ward objective function method. The application of the Forgy method follows in Section 2.

Although Sherman's factor analysis replication employs 117 medical institutions, 99 have been retained for cluster analysis. Data for the other 18 schools is missing for four (18%) or more of the 23 original vairables. Of the 99 medical schools used in this study, 83 are in common with the 94 used by RAND.

The particular variation of Ward's technique used in this analysis (Wishart, 1969) requires a symetric matrix of euclidean distances. In other words, a distance is computed for each pair of schools to reflect a composite of their differences on the six factor scores. All such combinations are stored in a 99 x 99 matrix and used as input for analysis by the Ward method.

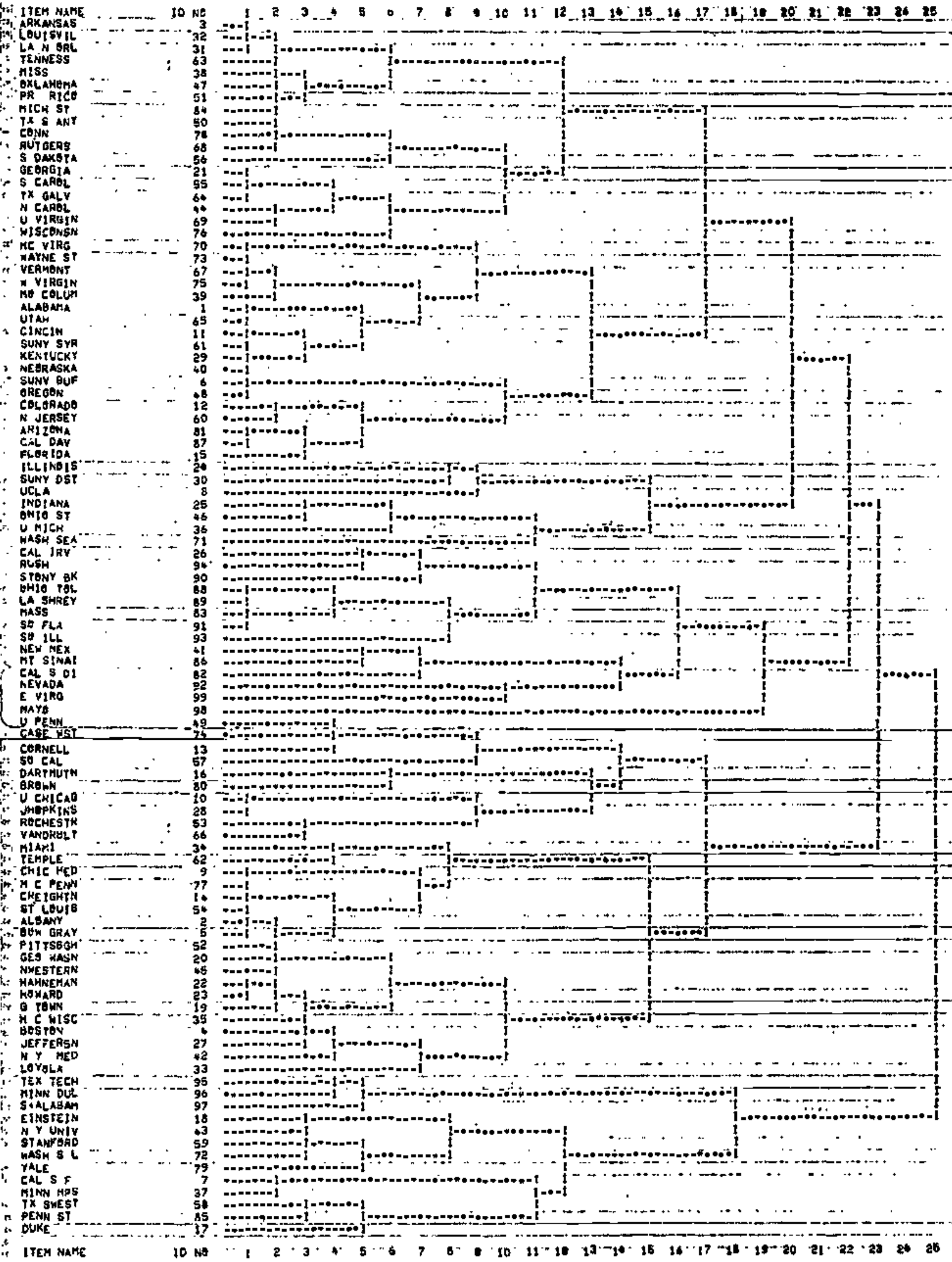
The result of the Ward analysis is provided in the form of a dendrogram in Figure 4.1. The dendrogram is condensed in order to reflect stages of the merge process rather than the full 98 individual (n-1) merges. These steps are equally incremented to preserve the actual error sum of squares needed to accomplish each merge. The 99 institutions analyzed are in the left column of the figure and the sequence number for 25 merge stages is shown across the top.

Figure 4.1 indicates that Arkansas, Louisville, and Louisiana-New Orleans merge at stage 1, that Tennessee joins this cluster at stage 2, and that this cluster does not admit any new members until stage 6 when a cluster consisting of Mississippi, Oklahoma, and Puerto Rico joins it.

On a broad perspective, the dendrogram reveals an elongated pattern of cluster growth. Merges between schools appear to occur fairly uniformly throughout the procedure; however, during the earlier stages they branch laterally for some distance before expanding vertically to admit larger numbers of institutions. This trend indicates that group structure is, on the whole, relatively introspective and confined to small memberships.

To illustrate this point, note that clusters forming up to stage 5 are numerous but generally contain only two to four members. In fact, at level 5 there are 44 clusters, or an average of only 2.25 schools per cluster. The largest cluster contains six schools, while a total of 17 schools have not as yet merged. If, for example, a single cluster is to be formed from the first 18 institutions listed in the dendrogram, the criterion level is found to be "12." The internal structure of this cluster in turn is made up of three subgroups of roughly

WARD GROUPING ON RAND REPL FACTORS *WT*



STOP 0

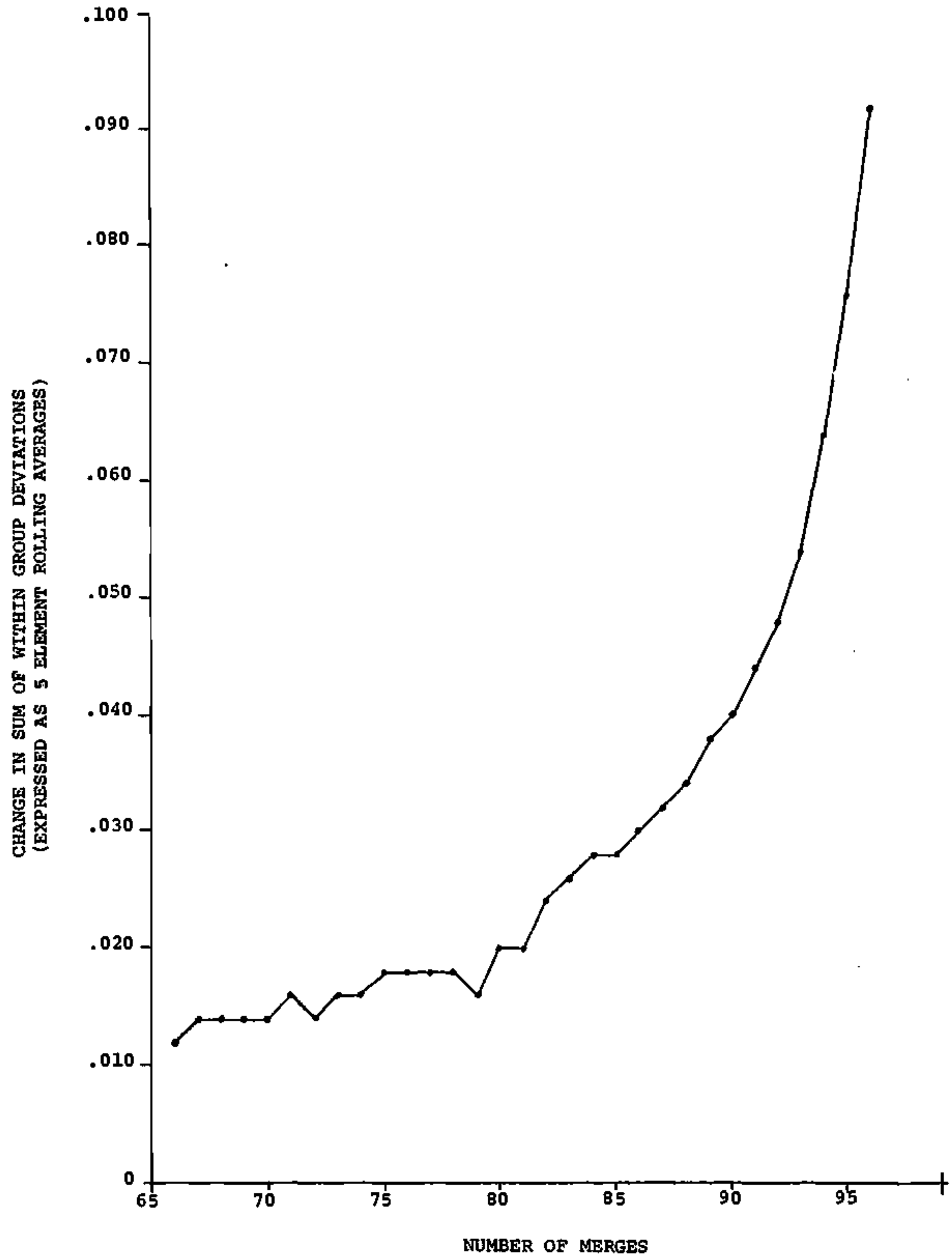
equal size. The criterion level has to be doubled, from level 6 to level 12, in order to tolerate the merging of these three subclusters.

In an earlier example, Figure 2.2, a plot of the number of clusters existing at any given stage against the critical distance has been suggested as a guide in selecting an optimal solution. It is perhaps more meaningful in the present context to plot the change in the sum of within-group deviations incurred at each merge. Additionally, to reduce the effects of local extremes, change in the sum of deviations is expressed as five element rolling averages. The resulting plot is given in Figure 4.2.

A plot of this kind often reveals disproportionate jumps transpiring between group deviations and the progression of merges. Figure 4.2, for instance, reveals perhaps two transitional points in the curve. The first point is at the eighty-first merge where 18 clusters have been formed: the second, at the ninetieth merge with nine clusters. These two transitional points suggest two optimal solutions. Schools grouped in the 18 cluster and 9 cluster solutions are given in Appendix B.

Ultimate determination of the "optimal" number of clusters is, of course, primarily subjective. Such determination must take into account the nature of the data analyzed, the methods used, and the particular goals of the analysis. The purpose of using hierarchical clustering

FIGURE 4.2


CHANGE IN SUM OF WITHIN GROUP DEVIATIONS
(EXPRESSED AS 5 ELEMENT ROLLING AVERAGES)
.100
.090
.080
.070
.060
.050
.040
.030
.020
.010
0
65
70
75
80
85
90
95
NUMBER OF MERGES

is first, to survey the structure of the data set and second, to provide seedpoints for the non-hierarchical analysis. Since the overall objective is to replicate the RAND Study, a judgment as to an optimal solution is predicated upon RAND's choice of ten clusters. Because there are 18 schools in the present study not considered in the RAND Study, it has been determined necessary to use a 16 cluster solution from the Ward analysis to generate seedpoints for the subsequent non-hierarchical application. The cluster seedpoints and cluster sizes for the 16 cluster solutions are given in Table 4.4.

2. Analysis by Forgy's K-Means Procedure

In the preceding application of the Ward method, a 99 x 99 similarity matrix has been used as input. In applying the Forgy method, however, the raw factor scores for each institution are employed. Again, these data consist of factor scores for six factors for each of 99 institutions.

The within-clusters summed deviations for the 16 cluster hierarchical solution is 109.15. The Forgy procedure, taking 5 iterations to converge, lowers this value to 97.61. A summary of the Forgy solution is given in Figure 4.3 in which the mean scores on each of the six factors for these 16 clusters are depicted. The names of schools belonging to each cluster is also given. Note that the line MD is the overall mean based on all 99 institutions.

TABLE 4.4

Assignment of Seedpoints*

CLUSTER NUMBER	NUMBER OF MEMBERS	SEEDPOINT COORDINATES					
		FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6
1	18	-.560	-.420	.520	.100	-.670	-.120
2	11	.150	-.510	.240	-.110	-.320	.360
3	7	.280	.430	-.400	-.290	-.530	.200
4	3	3.170	.470	.760	-.180	-1.040	-.610
5	4	.820	.800	1.080	-.700	-1.270	.890
6	8	.270	-.660	-1.670	-.180	-.580	-.060
7	3	.310	2.030	-1.590	-.360	-.690	.340
8	2	-2.020	.590	-1.300	-.720	.180	-1.090
9	1	4.660	-3.340	-3.780	-.500	.125	2.030
10	4	1.960	-.360	-.110	.190	1.250	.010
11	2	.080	-.270	-1.160	-.810	1.170	.900
12	4	.060	.870	-.320	.440	1.510	.380
13	9	-.490	-.430	.480	.110	.620	-.350
14	10	.290	.070	.180	-.480	.890	-.070
15	3	-1.370	-.420	-.840	1.280	-1.680	-.570
16	10	-.680	1.200	.610	.900	.280	-.380

* Seedpoint coordinates are derived by computing the mean score for each cluster on each of six factors.

FIGURE 4.3

Cluster Membership And Profile Summary

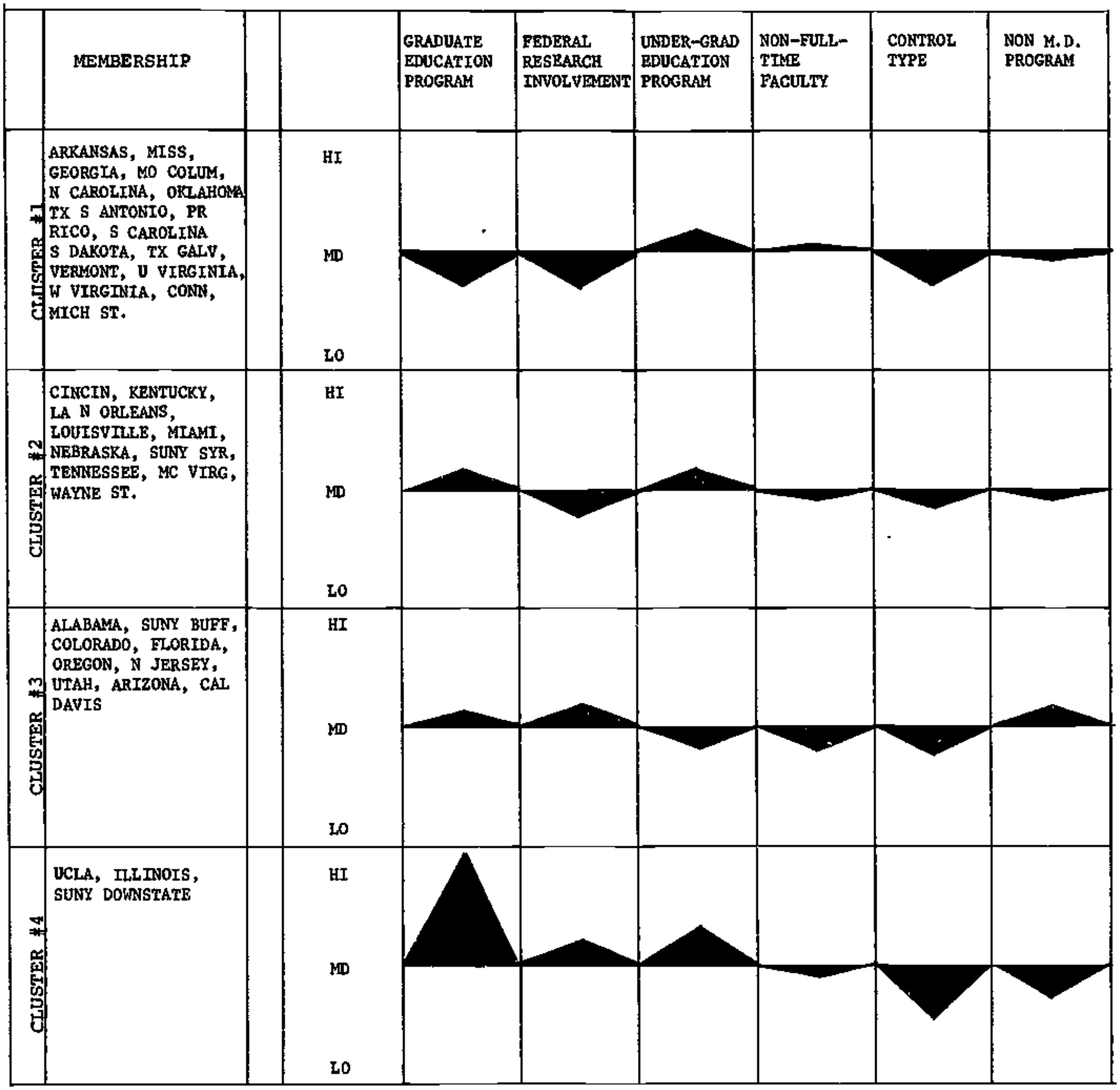


FIGURE 4.3 (Con't)

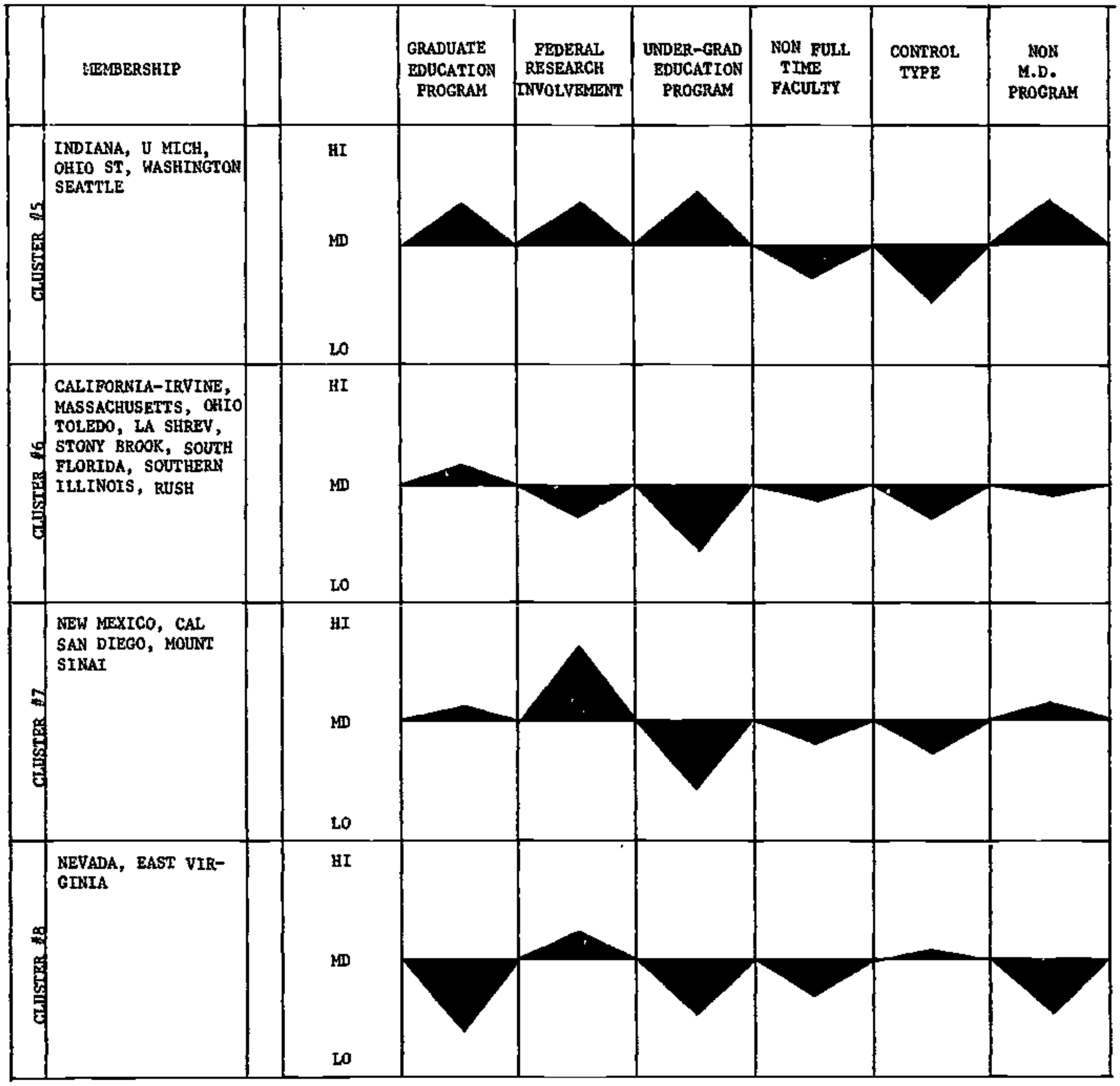


FIGURE 4.3 (Con't)

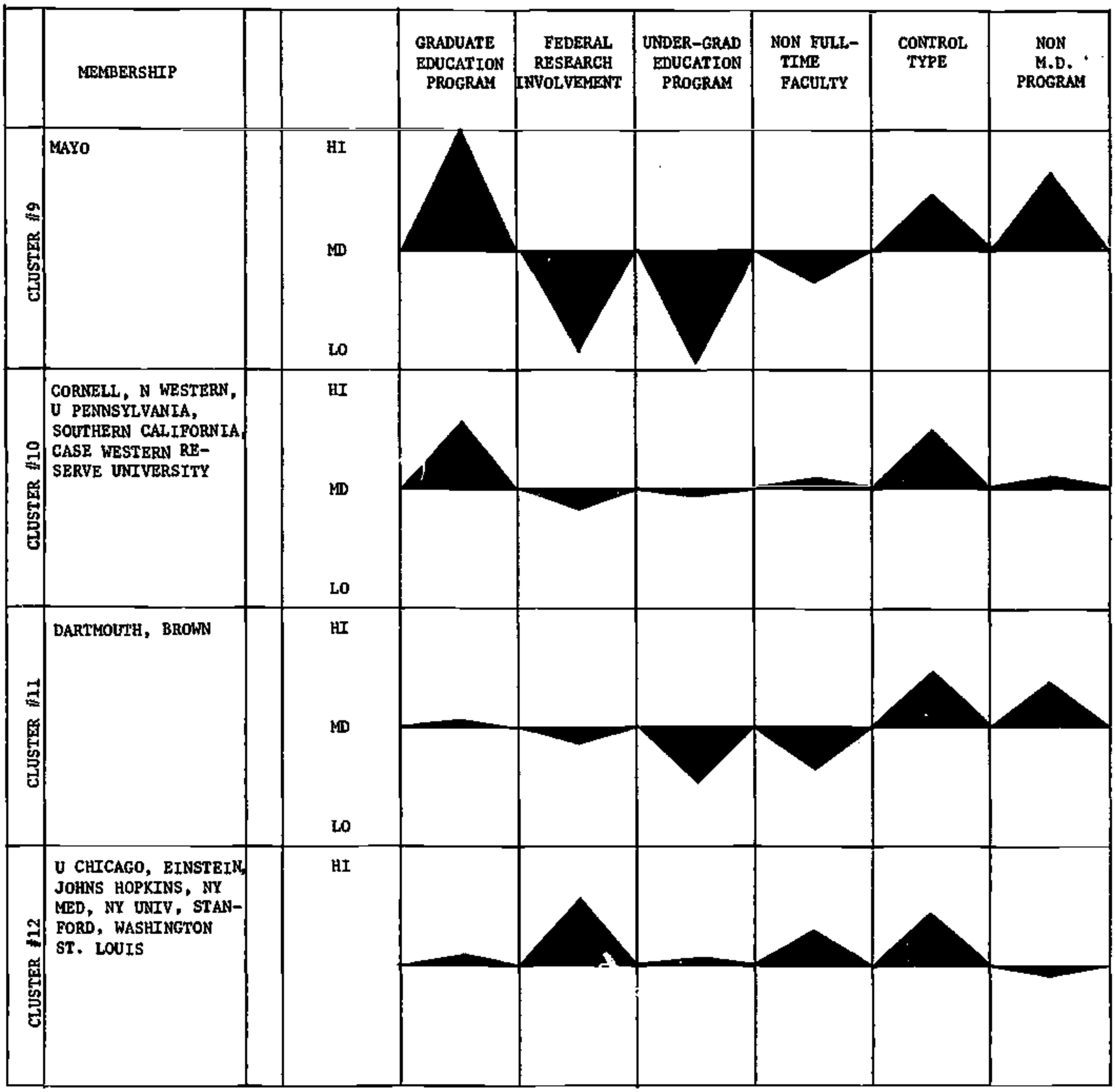
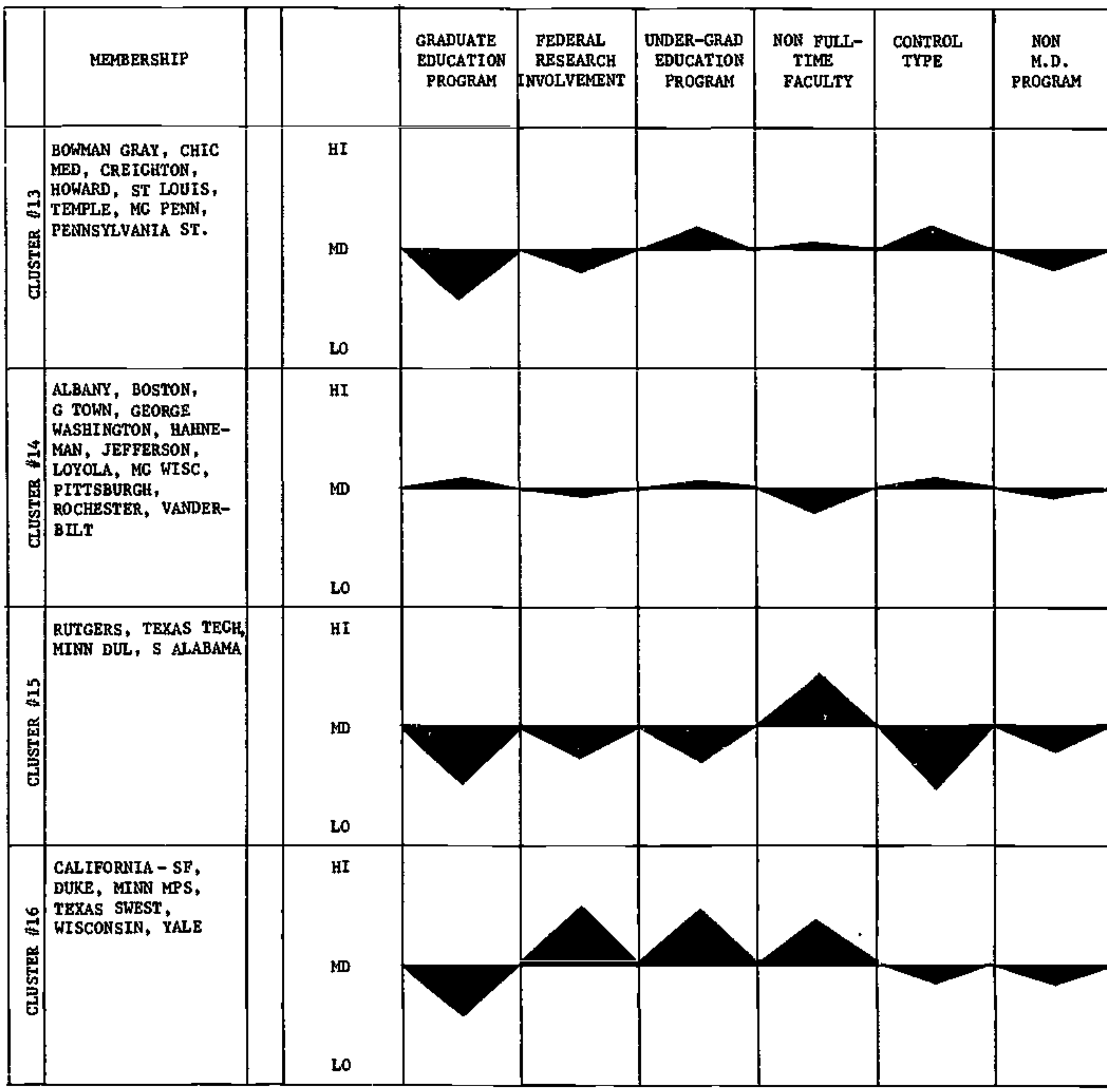


FIGURE 4.3 (Con't)



The first cluster contains 16 schools, all of which are public. As a group, these schools are somewhat below the mean in graduate program involvement and federal research involvement but generally have more undergraduate education programs. The profile of cluster 13 is virtually identical except with respect to control type. All the schools belonging to cluster 13 are private.

The profiles of clusters 2 and 14 are also similar to each other on all dimensions but control type. While the bulk of cluster 2 contains public schools, cluster 14 is entirely made up of private schools. Although cluster 14 is slightly closer to the mean on all six factors, the profiles on both indicate group structures characterized by higher emphasis on graduate and undergraduate education programs and a relatively low emphasis on federal research involvement, non-full-time faculty, and non-M.D. programs.

Clusters 3 and 7 have similar attributes except they differ in intensity. Both clusters include public schools that have larger than average graduate and non-M.D. programs and generally smaller undergraduate programs with less reliance on full-time faculty. However, cluster 7 has a much greater federal research commitment and a somewhat smaller undergraduate educational program.

Cluster 4 groups three public medical schools whose prominent feature is their large graduate educational program. Otherwise, these schools have a relatively high involvement in both federal research and undergraduate education, yet are somewhat below the mean in their reliance on non-full-time faculty and in non-M.D. programs offered. The four medical schools in cluster 5 have much the same attributes with two exceptions. First, the graduate education program is smaller and, second, the non-M.D. programs offered by these schools is comparatively high.

The differences between clusters 10 and 11 lie primarily in the intensity of scores on the six factors. First, the five schools in cluster 10 and two schools in 11 are all private. Both clusters have a lower than average commitment to federal research. Cluster 10 has a rather sizeable graduate education program with an average undergraduate education program and reliance on non-full-time faculty. The profile of cluster 11 on the three factors is, comparitively speaking, somewhat lower.

The profiles of the remaining five elusters are not directly comparable because their profiles approach varying extremes. Two of these clusters may well be considered outlying groups. Cluster 9, for instance, contains only Mayo. It can be readily seen from its profile that Mayo is quite unlike any other medical school in that it offers very large graduate and non-M.D. programs with an absence of involvement in federal research and undergraduate programs.

The two medical schools grouped in cluster 8, Nevada and Eastern Virginia, are new schools and as one would expect fall into lower extremes. In terms of research involvement, however, cluster 8 is slightly above the mean.

Cluster 15 is also made up of new medical schools. As a group, they differ from cluster 8 in their reliance on non-full-time faculty.

Cluster 12 consists of a group of all private schools distinguishable by heavy research involvement and reliance on non-full-time faculty. In all other regards, this cluster exhibits average characteristics.

A group of six public schools form cluster 6. These schools, besides having relatively extensive graduate programs, are otherwise on the lower extreme of the spectrum of variables.

Finally, the profile of cluster 16, which includes a mixture of public and private schools, indicates a lesser emphasis on graduate and non-M.D. educational programs, though the emphasis on undergraduate programs is substantial. Furthermore, the schools do have a significant involvement in federal research and reliance on non-full-time faculty.

D. Comparison of Results

A relatively simple way of comparing group membership between two sets of clustering results is to equate the number of matches and mismatches in tabular form, such as shown in Table 4.5. In the left hand column may be found the cluster identification numbers assigned by RAND to its resulting clusters. The cluster identification numbers shown across the top of the table are those evolving from the Forgy application to the replicated factors. For ease of interpretation, columns have been rearranged so as to reflect results of the two studies in corresponding order. In this way, clusters that are most highly associated in terms of membership fall roughly into cells along the diagonal.

Of those medical schools assigned by RAND to cluster 1, for example, consistency with the replicated cluster findings is minimal. The fifteen members of RAND's cluster 1 occur in small groups ranging across half the spectrum of the clusters derived from the replicated version. With few exceptions, much the same lack of conformity exists throughout. Notably, the exceptions are RAND's cluster 5, and to a lesser extent, 8 and 9. Members belonging to clusters 8 and 9 account for the bulk of private schools and differ mainly in regard to relative wealth and research orientation. The characteristics of the membership of cluster 5 are less obvious. Generally, the medical schools belonging to this cluster are public institutions with relatively low federal research involvement and graduate program emphasis.

TABLE 4.5

AAMC Cluster Numbers
As Assigned to Forgy Results

RAND CLUSTER NUMBERS

	5	6	2	3	4	10	1	7	13	12	14	16	8	9	11	15	Outlier
1	2	2	3	3			2					2					1
2				1			1	1									
3	1		2		2	1					3						2
4	1				1	3				2		1					2
5			2				9					1					1
6		1		1				1									
7			3	2		1	2		3		1						
8										4	2	2					2
9									5	1	5						2
10				2				1									1

As stated earlier, the number of clusters formed in the current analysis has been increased beyond RAND's ten to 16 in order to provide for the effects of imposing a fixed number of clusters on differing populations. Specifically, the concern has been to allocate a sufficient number of clusters to enable medical schools not in the RAND Study to develop into groups external to RAND's ten. As such, the several new medical schools forming the replicated clusters 8, 9, 11 and 15 are completely absent from the RAND analysis. The schools belonging to these four clusters then are separated out, thereby making the respective solutions more readily comparable. As a further note, no apparent combination of the 12 remaining replicated clusters enhances the over all conformity with the RAND clusters without, in turn, obscuring other equally important group distinctions.

Chapter V

Discussion and Conclusion

The present paper reports on two aspects of an attempt to utilize rather sophisticated statistical procedures to shed light on the "structure" of medical education in the U.S. The first aspect is primarily methodological, and it involves the descriptions in Chapter II and Chapter III regarding multivariate cluster analysis procedures and manipulation of available data into formats amenable to analysis by such procedures. The second aspect is primarily substantive, and it involves the description in Chapter IV regarding the generation of a classification of medical schools based on the methods and data developed, and comparison of these results to early efforts.

The discussion of the results reported, then, logically falls into two areas: methodological and substantive. The methodological aspect involves two components: choice of analysis procedures and constructing of a researchable data set. The present chapter is organized along these lines, with a final section devoted to conclusions that may be drawn from the report.

A. Methodological Considerations

1. Cluster Analysis Procedures

The cluster analysis procedures chosen for the present work are relatively new techniques in any data analyst's toolbox. As such, not a great deal is known about differences to be expected when one set of methods is used to contrast to some other set of methods. The lack of "maturity" of these methods contributes to the generally recognized concensus that use of such methods and interpretation of results is still more art than science.

Statistically based clustering procedures first received wide scale attention from applied research methods scholars in the mid- and late 1960's. A number of procedures predate this time frame, but these generally did not receive either wide scale attention nor use. With the wide availability of high speed electronic computers for scientific applications in the mid-60's, however, iterative approximation procedures became feasible, and as a result a wide number of statistically based cluster analysis procedures were suggested. Unfortunately but not atypically, at this stage of the game little is known about the effects of using particular methods in contrast to other methods.

The state of the art of statistically based clustering procedures may be contrasted to the state of the art of factor analysis. Much of the conceptual work on factor analysis procedures was done in the 1935 to 1955 time frame. In the 1960's, differences that could be expected in application of one procedure vs. another were delineated, and several widely acdepted "standard" procedures were adopted for general use, at least as an initial step, by applied researchers. The state of the art in 1975 of statistically based clustering procedures is not unlike the state of the art of factor analysis in the mid-1950's, one where many procedures have been suggested in the literature with little guidance available to the applied researcher as to which technique is best or at least most commonly used for various types of applications.

A very legitimate question to ask regarding the results in Chapter IV, and in particular regarding the lack of comparability of the results from the present study to the results of the RAND study, is that of "How much did the use of differing cluster analytic methodologies, albeit from the same framework, contribute to the lack of comparability?" Given the state of the art of cluster analysis, the answer must be "We don't know."

One hypothesis that has intuitive appeal and also a wide acceptance among scholars dealing with statistically based clustering procedures is that data sets having clear and unambiguous clusters of objects will be resolved appropriately by most of the techniques suggested. Data sets having unclear or ambiguous cluster structure are likely to yield differing results upon application of differing procedures. Such data sets are likely to yield widely variant results even upon application of very slight variants of one given procedure. Accepting this sort of hypothesis, and overlooking differences in results potentially due to data and substantive factors discussed below, one conclusion that might be drawn from the present study is that medical schools, as described by a wide range of institutional variables, present a sufficiently unclear and ambiguous set of objects such to resist meaningful or consistent resolution into groups via statistically based clustering procedures.

2. Construction of a Researchable Data Set

There are two elements in this aspect of the study that require some discussion. The first aspect is the availability of data and the selection of available data. Despite the abundance of data elements available in the IPS, there is little guidance as to which elements are most important or potentially most important in the present type of analysis. Important data elements may be missing from the available set; likewise, elements may be available but due to lack of knowledge, experience and/or previous research not selected for analysis. Continued efforts in the analysis of such data by procedures similar to the ones described in Chapter II are indicated as the only way to settle upon "key" data elements useful in any categorization of medical schools.

Second, the data preparation aspects necessarily involve arbitrary decisions. Scaling of data elements (i.e., standardization, transformations, etc.), and handling of missing data elements are of primary concern here. It is possible that further experience with the data and the methods will lead to variations in the data preparation procedures that substantively affect results. Again, only further efforts will shed light on the importance of these procedures.

B. Substantive Considerations

The results presented in Chapter IV indicate quite clearly that the classification of medical schools obtained by the RAND study was not replicated by the present study. The lack of replication may have been caused by a number of factors or combination of factors. Among potential explanatory factors are the methodological considerations discussed above, both procedural differences and data manipulative differences. The lack of comparability may also be due to one or more of a variety of substantive considerations.

First, the data used by RAND was 1969-70 data, whereas the data used by the present study was 1973-74 data. It is possible that in this four year period, schools changed their profiles, as reflected by the data analyzed, sufficiently to drastically change meaningful grouping of schools. In other words, it is possible that the classification structure presented by the RAND study was a "best" resolution for 1969-70 data and that the classification structure presented in the present paper is the "best" resolution based on 1973-74 data. Under this possible explanation, the differences are due to real changes that took place between 1969-70 and 1973-74, and these changes are reflected in the differing classificatory structures.

Second, the "quality" of the data analyzed by RAND and the "quality" of the data analyzed in the present study may be a factor in the lack of comparability of results. Much of the data analyzed by RAND was obtained from the AAMC. Over the past five years, the AAMC has conducted numerous activities to improve the quality and comparability of data collected from its constituency. It is felt that the quality, comparability, integrity and completeness of data collected has improved significantly during this time frame. Thus, the differences in the quality of the data analyzed by RAND and the quality of the data analyzed in the present effort may account at least in part for the lack of comparability of results.

Third, the measurement level of the data analyzed was different for many variables for the two efforts. Some of the data supplied by AAMC to RAND was of a sensitive nature for each school; thus to protect confidentiality, the data transmitted to RAND was converted to decile scores. The data analyzed by the present study was not subject to such a constraint. Such differences in measurement level for the data analyzed may again account at least in part for the lack of comparability of results.

Fourth, the set of variables analyzed by the two studies was not a complete match. Only 23 of the 31 variables analyzed by RAND were available, and even some of these were only approximations to the RAND variables. Even though the factor structure of the 23 variables seemed to be comparable to the factor structure of the 31 variables analyzed by RAND, it is possible that the factor scores were substantively different, due to the lack of variable set match.

Finally, the sets of schools analyzed by each study were not completely comparable. RAND analyzed data from 94 schools; the present study analyzed data from 99 institutions; there were only 83 schools common to both analyses. The differences in the analysis samples may have contributed to alterations in the measurement space sufficient to cause at least in part the non-comparability of results.

C. Conclusion

A number of conclusions may be drawn from the work presented in this report.

First, it should be concluded that classification methods based on statistically based cluster analysis methods have been developed and implemented for use on institutionally descriptive data stored in the AAMC's Institutional Profile System. This conclusion is directly relevant to the tasks to be accomplished by AAMC under contract.

Second, it also should be concluded that data from the AAMC's Institutional Profile System were extracted, massaged, and analyzed via the clustering procedures. This conclusion is again relevant to fulfillment of the contract.

Third, it may be concluded that the present study did not find evidence for the replicability of the results of the RAND study. In particular, the 10 clusters of medical institutions found by RAND were not found in the present study. A variety of factors were discussed that may have contributed to this conclusion, including a substantial number of both methodological and data differences between the studies.

Finally, it may be concluded that, at least for the present, categorization of medical schools via procedures accounting for multiple measures simultaneously does not yield clear and unambiguous results. Such a conclusion must be drawn given the lack of comparability of the present results and the RAND results.

This last conclusion should not be taken as a suggestion that such analyses are not useful; rather, it should be taken as an indication that the picture presented by data institutionally descriptive of U.S. medical schools is a highly complex one, one that despite the perceived need is not easily structured into a reasonably small number of groups of institutions. Now that such methods have been developed and are available for future utilization, it is logical to obtain further results to either substantiate or reject the conclusion. Further results are also needed to test the hypothesis that clear, unambiguous categorizations may be made based on analysis of subsets of variables from intuitively or empirically related domains.

BIBLIOGRAPHY

Anderberg, M.R. Cluster Analysis for Applications. New York: Academic Press, 1973.

Bailey, K.D. "Cluster Analysis." Sociological Methodology 1975. David Hess (Ed.) San Francisco: Jossey-Bass, 1974.

Forgy, E.W. Cluster Analysis of Multivariate Data: Efficiency Versus Interpretability of Classifications. Biometric Society Meetings, Riverside, CA (Abstract in Biometrics 21:3,768), 1965.

Gower, J.C. "A Comparison of Some Methods of Cluster Analysis." Biometrics, 23:628-637, Dec., 1967.

Greenhouse, S.W. "On the Meaning of Discrimination, Classification, Mixture, and Clustering in Statistics." The Role and Methodology of Classification in Psychiatry and Psychopathology. M.M. Katz, J.O. Cole, and W.E. Barton, (Ed.). Chevy Chase, MD. U.S. Dept. of HEW, 1967.

Hartigan, J.A. Clustering Algorithms. New York: Wiley and Sons, 1975.

Jancey, R.C. "Multidimensional Group Analysis." Austrailian Journal of Botany, 14:127-130, 1966.

Jardine, N. and Sibson, R. "The Construction of Hierarchic and Non-hierarchic Classifications." Computer Journal, 177-184, Aug., 1968.

Johnson, S.C. "Hierarchical Clustering Schemes." Psychometrika, 32:241-254, Sept., 1967.

Keeler, E., Koehler, J.E., Lee, C., and Williams, Jr., A.P. Finding Representative Academic Health Centers. A Working Note prepared for NIH/HEW. Santa Monica, CA: Rand Corp., 1972.

Lance, G.N. and Williams, W.T. "Computer Programs for Hierarchical Polythetic Classification." Computer Journal, 9:60-64, March, 1966.

_______ "A General Theory of Classificatory Sorting Strategies: Hierarchical Systems." Computer Journal, 9:373-380, Feb., 1967.

McRae, D.J. "MIKCA: A FORTRAN IV Iterative K-Means Cluster Analysis Program" Behavioral Science, 16:423-424, 1971.

McQuitty, L.L. "Hierarchical Linkage Analysis for the Isolation of Types." Educational and Psychological Measurement, 20:55-67, Spring, 1960.

Mulaik, S.A. The Foundations of Factor Analysis. New York: McGraw-Hill, 1972.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., and Bent, D.H. Statistical Package for the Social Sciences. Second Edition. New York: McGraw-Hill, 1975.

Otis, G.D., Graham, J.R., and Thacher, L. "Typological Analysis of U.S. Medical Schools." Journal of Medical Education, 50:328-338, 1975.

Rodgers, S.A. and Elton, C.F. "An Analysis of the Environment of Medical Schools." Research in Higher Education, 2:239-249, Sept., 1974.

Rummel, R.J. "Understanding Factor Analysis." Journal of Conflict Resolution. 2:444-480, March, 1967.

Sedlacek, W.E. Variables Related to Increase in Medical School Class Size. Washington, D.C.: Association of Americal Medical Colleges, 1975.

Sells, S.B. "Toward a Taxonomy of Organizations." New Perspectives in Organizational Research. W.H. Cooper, H.J. Leavitt, and M.W. Shelly (Eds.) New York: John Wiley, 1964.

Shaw, K.E. "An Application of Cluster Analysis to the Study of Organization in a College of Education." Higher Education, 2:343-356, Aug., 1973.

Sherman, C.S. Study of Medical Education: Interrelationships Between Faculty, Curriculum, Student and Institution Variables. Washington, D.C.: Association of American Medical Colleges, 1975.

Tryon, R. and Bailey, D.E. Cluster Analysis. New York: McGraw-Hill, 1970.

Ward, Jr., J.H. "Hierarchical Grouping to Optimise an Objective Function." Journal of American Statistical Association, 58:236-244, 1963.

Wishart, D. "An Algorithm for Hierarchical Classification." Biometrics, 22:165-170, 1969.

ABBREVIATIONS

ACADM	ACADEMIC
ADMISS	ADMISSIONS
ADMN & GEN	ADMINISTRATIVE & GENERAL
ADV DEGREE	ADVANCED DEGREE
ADVIS PROG	ADVISORY PROGRAM
AFFIL	AFFILIATED
AM	AMERICAN
AMBUL	AMBULATORY
AMT	AMOUNT
ANESTH	ANESTHESIOLOGY
APPL	APPLICANT, APPLICATION
ASSOC PROF	ASSOCIATE PROFESSOR
ASSOC PROF MD	ASSOCIATE PROFESSOR OF MEDICINE
ASSTD	ASSISTED
AV	AVERAGE
BACH	BACHELORS DEGREE
BAS SCI	BASIC SCIENCE
BEHAV OBSS PUBLSHD	BEHAVIORAL OBJECTIVES PUBLISHED
BLDG	BUILDING
CL SCI	CLINICAL SCIENCE
CONSTR	CONSTRUCTION
CURR	CURRICULUM
DEPT	DEPARTMENT
DEV	DEVELOPMENT
DOC	DOCTORATE
DOC CAND	DOCTORAL CANDIDATE
DOC CONFRD	DOCTORALS CONFERRED
ED	EDUCATION
ENDOW	ENDOWMENTS
ENTERING	ENTERING STUDENTS
EQUIP	EQUIPMENT
EXPD	EXPENDITURES
FAC	FACULTY
FED	FEDERAL
FMS	FOREIGN MEDICAL STUDENTS
FT FAC	FULL-TIME FACULTY
GPA	GRADE POINT AVERAGE
GRAD	GRADUATION
GRTS	GRANTS

HLTH	HEALTH
HMO	HEALTH MAINTENANCE ORGANIZATIONS
HOSPS	HOSPITALS
HS SR	HIGH SCHOOL SENIOR
INDUS	INDUSTRY
INNOVATN	INNOVATION
INSTR	INSTRUCTOR, INSTRUCTED
INSTR & DEPT RESRCH	INSTRUCTION & DEPARTMENTAL RESEARCH
LOC	LOCAL
MANGMT	MANAGEMENT
MAS	MASTERS DEGREE
MC	MEDICAL COLLEGE
MCAT SCORE GEN	MCAT SCORE GENERAL KNOWLEDGE
MCAT SCORE SCI	MCAT SCORE SCIENCE
MCAT SCORE VER	MCAT SCORE VERBAL
MCAT SCORE QUAN	MCAT SCORE QUANTITATIVE
MD	MEDICAL
NATL BDS	NATIONAL BOARDS
NEED & RECVD AID	NEEDED & RECEIVED AID
NON-GOVT	NON-GOVERNMENT
PCT	PERCENT
PHYS ASST	PHYSICIAN'S ASSISTANT
POP	POPULATION
PRIM CARE	PRIMARY CARE
PRIV	PRIVATE
PROF	FULL PROFESSOR
PROF MD	PROFESSOR OF MEDICINE
PROG	PROGRAM
PROJ	PROJECT
PROJTD	PROJECTED
PT FAC	PART-TIME FACULTY
RECVD	RECEIVED
REG OP COSTS	REGULAR OPERATING COSTS
REQ AID	REQUESTED AID
REQ & RECVD AID	REQUESTED & RECEIVED AID
RESDNTS	RESIDENTS
RESRCH	RESEARCH
REV	REVENUES
REV CAREER	REVIEW CAREER

SCH	SCHOOL
SELECTD	SELECTED
SERV	SERVICE
SMSA	STANDARD METROPOLITAN STATISTICAL AREA
SPONS	SPONSORED
ST	STATE
STUDENT EQUIV	STUDENT EQUIVALENT
TCH-TRN	TEACHING-TRAINING
TOT	TOTAL
TRANS STUDENTS	TRANSFERRED STUDENTS
TUIT & EXPEN	TUITION & EXPENSES
UNIV	UNIVERSITY
UNRESTR	UNRESTRICTED
VOL	VOLUMES
VOL FAC	VOLUNTARY FACULTY
WITHDRL	WITHDRAWALS
YR	YEAR

APPENDIX A

VARIABLES LIST FOR CLASSIFICATION
OF MEDICAL INSTITUTIONS STUDY
AND INTERRELATIONSHIPS STUDY

	INSTITUTION	IPS SOURCE VARIABLES	MATHEMATICAL TRANSFORMATIONS
	*** GENERAL CHARACTERISTICS***		
V1000	MC-IDENTIFICATION CODE	IPS	
V1010	STATE MC LOCATED	IPS	
V1020	REGION MC LOCATED	IPS	
V1030	CONTROL TYPE (LOW=PUBLIC HIGH=PRIVATE)	IPS	
V1040	YEAR FOUNDED	3064	
V1045	AGE OF INSTITUTION	YR 1974-3064	
V1050	2 OR 4 YR SCH	3066	
V1060	ACCREDITATION	3065	
V1070	MC TYPE & HOSPITAL	2847	
V1071	UNIV AFFIL HOSPITAL	2847	
V1072	UNIV OR ANY AFFIL HOSPITAL	2847	
V1080	TOT BEDS AFFIL HOSPITAL	American Hospital Association,Curriculum Directory	
V1085	RATIO AFFIL HOSP BEDS TO MD STUDENTS	American Hospital Association,Curriculum Directory	
V1090	NUMBER OF DEANS APPNTD 60-74	Department of Institutional Development	
	*** OEMOGRAPHIC ***		
V1100	MC LOCATION-SMSA POP 71	0366	
V1110	MC LOCATION-IMMEDIATE LOCATION POP 71	0367	
V1120	MC LOCATION-IMMEDIATE LOCATION POP-DENSITY 71	0368	
V1130	MC LOCATION-SMSA POP-PCT NON-WHITE	0369	
V1140	SMSA POP PER MD STUDENT	0366/1391	
	*** LIBRARY ***		
V1200	MC LIBRARIES-TOT VOL	2223	
V1210	MC LIBRARIES-ACQUISITIONS	2224	
V1220	MC LIBRARIES-TOT SERIAL TITLES RECVD	2225	
	FINANCES (ACADEMIC YR 72-73)		
	*** REVENUES ***		
--TOTALS BY SOURCE--			
V2000	MC REV-TOT ALL SOURCES	1120	/1000
V2010	MC REV-TOT FED SOURCES	3129	/1000
V2017	PCT OF MC REV FROM FED SOURCES + INDIRECT COST RECOVERY	(1112 + 3129)/ 1120	x100
--TOTALS BY SOURCE (UNRESTR)--			
V2100	MC REV-TOT UNRESTR PROFESSIONAL FEES,MD SERV PLANS	1118	/1000
V2110	MC REV-TOT UNRESTR ENDOW & GIFTS	1093	/1000
V2115	PCT OF TOT MC REV FROM UNRESTR ENDOW & GIFTS	(1094+1098/1120)	x100

V2120	MC REV-TOT UNRESTR STUDENT TUITION & FEES	1084	/1000
V2125	PCT OF TOT MC REV FROM UNRESTR STUDENT TUITION & FEES	(1086/1120-1102-110,-1111-1'14)	x100
V2130	MC REV-TOT UNRESTR FED, ST, LOC SOURCES	1092/1000	
V2140	MC REV-TOT UNRESTR GIFTS BUSINESS & INDUS	1096	/1000
V2145	PCT OF TOT MC PEV FROM UNRESTR GIFTS BUSINESS & INDUS	(1096/1098)	x100
V2150	MC REV-TOT UNRESTR GIFTS FOUNDATION	1095	/1000
V2155	PCT OF TOT MC REV FROM UNRESTR GIFTS FOUNDATIONS	(1095/1098)	x100
V2160	MC REV-TOT UNRESTR GIFTS ALUMNI	1094	/1000
V2165	PCT OF TOT REV FROM UNRESTR GIFTS ALUMNI	(1094/1098)	x100
V2170	MC REV-TOT GIFTS	1098	/1000
--RECOVERY OF INDIRECT COSTS OF SPONS PROGS--			
V2200	MC REV-TOT INDIRECT COSTS RECOVERY	1115	/1000
V2210	MC REV-INDIRECT COSTS RECOVERY NON-GOVT	1114	/1000
V2220	MC REV-INDIRECT COSTS RECOVERY FED PROG	1112	/1000
--SPONSORED TOTALS BY SOURCE--			
V2300	MC REV-TOT FED SPONS PROG	3129	/1000
V2310	MC REV-TOT MULTI & SERV SPONS PROG	1111	/1000
--SPONSORED RESEARCH BY SOURCE--			
V2400	MC REV-TOT SPONS RESRCH	1102	/1000
V2405	PCT OF TOT MC REV FOR SPONS RESRCH	(1102/1120)	x100
V2410	MC REV-TOT FED SPONS RESRCH	1099	/1000
V2415	PCT OF TOT SPONS RESRCH FROM FED	(1099/1102)	x100
V2420	MC REV-TOT ST, LOC SPONS RESRCH	1100	/1000
V2425	PCT OF TOT SPONS RESRCH FROM ST,LOC	(1100/1102)	x100
V2430	MC REV-TOT NON-GOVT SPONS RESRCH	1101	/1000
V2435	PCT OF TOT SPONS RESRCH FROM NON-GOVT	(1100+1101/1102)	x100
--SPONSORED TCH-TRN BY SOURCE--			
V2500	MC REV-TOT SPONS TCH-TRN	1107	/1000
V2505	PCT OF TOT MC REV FROM SPONS TCH-TRN	(1107/1120)	x100
V2510	MC REV-TOT FED SPONS TCH-TRN	1104	/1000
V2515	PCT OF TOT SPONS TCH-TRN FROM FED	(1104/1107)	x100
V2520	MC REV-TOT ST,LOC SPONS TCH-TRN	1105	/1000
V2525	PCT OF TOT SPONS TCH-TRN FROM ST,LOC	(1105/1107)	x100
V2530	MC REV-TOT NON-GOVT SPONS TCH-TRN	1106	/1000
V2535	PCT OF TOT SPONS TCH-TRN FROM NON-GOVT	(1105+1106/1107)	x100
--	*** EXPENDITURES ***		
--TOTALS BY FUNCTIONAL CATEGORY (UNRESTR)--			
V2600	MC EXPD-TOT UNRESTR	1137	/1000
V2610	MC EXPD-TOT UNRESTR ADMN & GEN	1136	/1000
V2615	PCT OF TOT UNRESTR MC EXPD FOR ADMN & GEN	(1136/1137)	x100
V2620	MC EXPD-TOT UNRESTR ACADM SALARY, FEES TOT ACTUAL	1251	/1000

V2625	PCT OF TOT UNRESTR MC EXPD FOR ACADM SALARY,FEES	(1251/1137)	x100
V2630	MC EXPD-TOT UNRESTR INSTR & DEPT RESRCH	1124	/1000
V2635	PCT OF TOT UNRESTR MC EXPD FOR INSTR & DEPT RESRCH	(1124/1137)	x100
V2640	MC EXPD-TOT UNRESTR PUBLIC SERV	1130	

--EXPENDITURES PER STUDENT & STAFF--

V2700	INSTR & DEPT RESRCH EXPD PER STUDENT	1126/(1257+0551+0550+3130+3137+1559)	
V2710	INSTR & DEPT RESRCH EXPD PER FAC	1126/3127	
V2720	MC EXPD-TOT UNRESTR PER MD STUDENT	1137/1257	
V2730	MC EXPD-TOT UNRESTR PER FT FAC	1137/3127	
V2740	SPECIAL PROJ $ PER MD STUDENT 72-73	1205/1257	
V2750	TOT MC EXPD PER TOTAL STUDENTS	1137/(1391+1559+3130+1549+1548)	

--SPONSORED EXPENDITURES--

V2800	MC EXPD-TOT SPONS RESRCH	1126	/1000
V2805	PCT OF TOT MC EXPD FOR SPONS RESRCH	(1126/1137)	x100
V2810	MC EXPD-TOT SPONS TCH-TRN	1128	/1000
V2815	PCT OF TOT MC EXPD FOR SPONS TCH-TRN	(1128/1137)	x100
V2820	PCT SPONS FAC SALARY FROM FED S 72-73	(1162/1168)	x100
V2830	MC EXPD-PCT SPONS PROG EXPD OF TOT	(1159/1137)	x100
V2840	MC EXPD-TOT SPONS PROGS--ALL TYPES	1159	/1000

*** NIH AWARDS ***

V2900	NIH AWARDS-PROG+PROJ & CENTER GRTS $1000	1120	/1000
V2910	NIH AWARDS-RESRCH GRTS $1000 67-68	2249	/1000
V2920	NIH AWARDS-RESRCH GRTS $1000 68-69	2250	/1000
V2930	NIH AWARDS-RESRCH GRTS $1000 72-73	2254	/1000
V2940	NIH AWARDS-RESRCH GRTS $1000 73-74	2255	/1000
V2950	NIH AWARDS PCT CHANGE	(2250-2249/2249)+(2254-2250/2250)+(2255-2254/2254/3)	x100
V2951	NIH RESRCH $ PCT CHANGE	(2254+2255)-(2249+2250)/(2249+2250)	x100

*** CONTRUCTION***

--FUNDS BY SOURCE--

V3000	CONSTR FUNDS-TOT FED	1937	/1000
V3005	PCT OF TOT CONSTR FUNDS FROM FED	(1937/1935)	x100
V3010	CONSTR FUNDS-TOT ST	1938	/1000
V3015	PCT OF TOT CONSTR FUNDS FROM ST	(1938/1935)	x100
V3020	CONSTR FUNDS-TOT PRIV GIFTS	1939	/1000
V3025	PCT OF TOT CONSTR FUNDS FROM PRIV GIFTS	(1939/1935)	x100
V3030	CONSTR FUNDS-TOT OTHER	1940	/1000
V3035	PCT OF TOT CONSTR FUNDS FROM OTHER	(1940/1935)	x100

--BUILDING COSTS--

V3100	BLDG CONSTR COSTS-TOT	1935	/1000
V3110	MOVABLE EQUIP CONSTR COSTS-TOT	1936	/1000

--BUILDING USE--

V3200	CONSTR BLDG USE-PCT FOR TCH	1941	/1000
V3210	CONSTR BLDG USE-PCT FOR RESRCH	1942	/1000
V3220	CONSTR BLDG USE-PCT FOR MD SERV	1943	/1000
V3230	CONSTR BLDG USE-PCT FOR OTHER	1944	/1000
	*** GENERAL ***		
V3300	PROFESSIONAL FEES RECVD PER CL SCI FAC	(1118/1030)	x100
V3310	MC LIBRARIES-BUDGET,BOOKS,PERIODICALS,BINDING	2218	
V3320	MC EXPEN-SPONS RESRCH PER FT FAC	1126/3127	
V3325	MC EXPEN-SPONS RESRCH PER MD STUDENT	1126/1257	
V3330	MC EXPEN-SPONS TCH-TRN PER MD STUDENT	1128/1257	
V3340	MC EXPEN-REG OP COSTS		
V3346	MC REV-TOT PER MD STUDENT	1120/1391	
V3350	SPONS PROG EXPD PER FT FAC	1159/3132	

ACADEMIC PROGRAM

*** GENERAL ***

V4000	OFFER COMBINED DOC+MD PROG 74-75	1321
V4010	USE NATL BDS PT 1-PROMOTION TEST 74-75	1359
V4020	USE NATL BDS PT 2-GRADUATION TEST 74-75	1362
V4030	MINIMUM MONTHS INSTR FOR MD DEGREE	2059
V4035	UNIT FOR RESRCH & DEV OF ED PROCESS	1378
V4040	MC PERMITS PASS-FAIL GRADING	1352
V4050	TYPE GRADING-HONORS,PASS,FAIL 74-75	1353
V4060	HLTH PRACTITIONER PROG-PHYS ASST 73	0387
V4070	HLTH PRACTITIONER PROG-NURSING 73	0388
V4080	HLTH PRACTITIONER PROG-MEDEX 73	0389
V4090	HLTH PRACTITIONER PROG-MIDWIFE NURSE 73	0390

*** CURRICULUM ***

V4100	CURR INNOVATN-AMBUL PRIM CARE PROG 74-75	1350
V4110	CURR INNOVATN-SPECLTY TRACKS 74-75	1351
V4120	CURR INNOVATN-CL APPL COMPUTERS 74-75	1343
V4130	CURR INNOVATN-COMPUTER ASSTD INSTR 74-75	1344
V4140	CURR ELECTIVES-HUMAN SEXUALITY 74-75	1332
V4150	CURR ELECTIVES-MD JURISPRUDENCE 74-75	1333
V4160	CURR ELECTIVES-NUTRITION 74-75	1334
V4170	CURR ELECTIVES-NON-WESTERN MEDICINE74-75	1335
V4180	CURR ELECTIVES-POP DYNAMICS 74-75	1336
V4190	CURR ELECTIVES-DRUG ABUSE 74-75	1337
V4200	CURR ELECTIVES-ALCOHOLISM 74-75	1338
V4210	CURR ELECTIVES-MD HYPNOSIS 74-75	1339
V4220	CURR ELECTIVES-ETHICAL PROBLEMS 74-75	1340
V4230	CURR ELECTIVES-HLTH CARE DELIVERY 74-75	1341
V4240	CURR-FAMILY MD PROG 74-75	2066

V4250	CURR-FAMILY MD GRAD PROG 73	0403	
V4260	CURR-PRIMARY CARE PROG 74-75	2071	
V4270	CURR-ACCELERTD PROG-MD DEGREE LESS THAN 6 YRS	1310	
V4280	CURR-RESRCH & DEV OF ED PROCESS 74-75	1378	
V4290	CURR-REQUIRED AMBUL CARE EXPERIENCE 73	0370	
V4300	CURR-PCT UNDERGRAD EXPERIENCE AMBUL CARE 73	0372	
V4310	CURR-PRIM CARE DEPT ENCOURAGE GENERALIST 73	0375	
V4320	CURR-TOT MD STUDENTS OPERATIONAL HMO 73	0381	
V4325	CURR-HLTH PRACTITIONER PROG 73		
V4330	CURR-EMERGENCY CARE PROG 73	0418	
V4340	CURR-PATIENT CARE PROG-ALCOHOLISM OR DRUG ABUSE73	0420	
V4350	CURR-HLTH CARE MANGMT PROG 73	0424	
V4360	STATEMNT OF BEHAV OBJS PUBLSHD	1374	

FACULTY

*** STAFF ***

--TOTAL TEACHING STAFF--

V5000	FT FAC-TOT ALL DEPT 72-73	3127	
V5010	FT FAC-TOT ALL DEPT 73-74	3132	
V5020	RATIO-FT FAC TO MD STUDENTS	1391/3132	
V5025	RATIO FT FAC TO TOTAL STUDENTS	3132/1391+1559+3130+1549+1548	
V5030	RATIO PT FAC TO FT FAC	(1734+1786)/3132	
V5040	RATIO VOL FAC TO FT FAC	(1768+1804)/3132	

--TOTALS BY MAJOR DISCIPLINE--

V5100	BAS SCI-TOT FT FAC	1662	
V5110	BAS SCI-TOT PT FAC	1734	
V5120	BAS SCI-TOT VOL FAC	1768	
V5130	CL SCI-TOT FT FAC 72-73	1030	
V5140	CL SCI-TOT FT FAC 73-74	1680	
V5150	CL SCI-TOT PT FAC	1786	
V5160	CL SCI-TOT VOL FAC	1804	
V5170	RATIO-BAS SCI FAC TO CLIN SCI FAC	1662/1680	

--TOTALS BY RANK--

V5200	PROF-TOT FT-CLI SCI	1680	
V5205	PROF-PCT FT-CLI SCI	(1680/1752)	x100
V5210	ASSOC PROF-TOT FT-CLI SCI	1698	
V5215	ASSOC PROF-PCT FT-CLI SCI	(1698/1752)	x100
V5220	ASST PROF-TOT FT-CLI SCI	1716	
V5225	ASST PROF-PCT FT-CLI SCI	(1716/1752)	x100
V5230	INSTR-TOT FT-CLI SCI	1734	
V5235	INSTR-PCT FT-CLI SCI	(1734/1752)	x100
V5240	PROF-TOT FT-BAS SCI	1630	
V5245	PROF-PCT FT-BAS SCI	(1630/1662)	x100
V5250	ASSOC PROF-TOT FT-BAS SCI	1638	
V5255	ASSOC PROF-PCT FT-BAS SCI	(1638/1662)	x100

V5260	ASST PROF-TOT FT-BAS SCI	1646	
V5265	ASST PROF-PCT FT-BAS SCI	(1646/1662)	x100
V5270	INSTR-TOT FT-BAS SCI	1654	
V5275	INSTR-PCT FT-BAS SCI	(1654/1662)	x100

--VACANCIES--

V5300	VACANCIES-FT FAC-CL SCI	1934
V5310	VACANCIES-FT FAC-BAS SCI	1844
V5320	PCT BUDGETED VACANCIES-CL SCI	1934/(1934+1752)

*** SALARY ***

--BASIC SCIENCE BY RANK--

V5400	AV TOT SALARY-PROF-BAS SCI 74-75	3579
V5410	AV TOT SALARY-ASSOC PROF-BAS SCI 74-75	3580
V5420	AV TOT SALARY-ASST PROF-BAS SCI 74-75	3581
V5430	AV TOT SALARY-INSTR-BAS SCI 74-75	3582

--CLINICAL SCIENCE BY RANK--

V5500	AV TOT SALARY-PROF-CL SCI 74-75	3584
V5510	AV TOT SALARY-ASSOC PROF-CL SCI 74-75	3586
V5520	AV TOT SALARY-ASST PROF-CL SCI 74-75	3587
V5530	AV TOT SALARY-INSTR-CL SCI 74-75	3588

--DEPARTMENT OF MEDICINE BY RANK--

V5540	AV TOT SALARY-PROF MD-CL SCI 74-75	3640
V5550	AV TOT SALARY-ASSOC PROF MD-CL SCI 74-75	3641
V5560	AV TOT SALARY-ASST PROF MD-CL SCI 74-75	3642
V5570	AV TOT SALARY-INSTR MD-CL SCI 74-75	3643

--ANESTHESIOLOGY BY RANK--

V5600	AV TOT SALARY-PROF-ANESTH 74-75	3620
V5610	AV TOT SALARY-ASSOC PROF-ANESTH 74-75	3621
V5620	AV TOT SALARY-ASST PROF-ANESTH 74-75	3622
V5630	AV TOT SALARY-INSTR-ANESTH 74-75	3623

STUDENT ADMISSIONS

*** ENROLLMENT ***

--STUDENT BODY TOTALS--

V6000	ENROLL-TOT STUDENTS	1391+1548+1549+3138+3130
V6010	TOT STUDENTS...ALL...INSTRUCTED AT MC	1391+1559+3130+1549+1548
V6020	ENROLL-TOT MD STUDENTS 73-74	1391

V6025	ENROLL-TOT MD STUDENTS 72-73	1257	
V6030	ENROLL-ACTUAL GROWTH RATE	(1391-1257)/1257	x100
V6040	ENROLL-TOT MD STUDENT EQUIV INSTR BY MD	1559	
V6050	ENROLL RATIO-MD STUDENTS EQUIV TO MD STUDENTS	1559/1391	
V6080	ENROLL RATIO-INTERNS & RESDNTS TO MD STUDENTS	(1549+1548)/1391	
V6090	ENROLL RATIO-INTERNS TO MD STUDENTS	1549/1391	
V6100	ENROLL RATIO-RESDNTS TO MD STUDENTS	1548/1391	
V6110	ENROLL-TOT FINAL YR STUDENTS-MAS & DOC CAND-BAS SCI	3130	
V6120	ENROLL-TOT FINAL YR STUDENTS-MAS & DOC CONFRD	3131	
V6130	ENROLL-TOT FINAL YR STUDENTS-NON-DEGREE CAND	3137	
V6140	ENROLL RATIO-MAS & DOC BAS SCI TO MD STUDENTS	3130/1391	
V6160	ENROLL RATIO-MAS & DOC CONFRD TO TOT ENROLL	3131/1391+1548+1549+3138+3130	

--IN STATE-OUT OF STATE STUDENTS--

V6200	ENROLL-TOT IN ST MD STUDENTS	1970	
V6210	ENROLL-TOT OUT ST MD STUDENTS	1971	
V6220	ENROLL RATIO-IN ST TO OUT ST MD STUDENTS	1970/1971	
V6230	PCT MD STUDENT FROM HOME STATE	1970/1391	

--STUDENTS PER FACULTY--

V6300	TOT RESDNTS INSTR BY MD FAC 72-73	0551	
V6310	TOT RESDNTS INSTR BY MD FAC 73-74	1549	
V6320	TOT INTERNS INSTR BY MD FAC 72-73	0550	
V6330	TOT INTERNS INSTR BY MD FAC 73-74	1548	

--PROJECTED ENROLLMENT--

V6400	PROJTD ENROLL-TOT FINAL YR MD STUDENTS 74-75	1620	
V6410	PROJTD ENROLL-TOT FINAL YR MD STUDENTS 75-76	1621	
V6420	PROJTD ENROLL-TOT FINAL YR MD STUDENTS 76-77	1622	
V6430	PROJTD ENROLL-PCT GROWTH MD STUDENTS 74-77		
V6440	PROJTD ENROLL-TOT 1ST YR MD STUDENTS 74-75	1610	
V6450	PROJTD ENROLL-TOT 1ST YR MD STUDENTS 75-76	1611	
V6460	PROJTD ENROLL-TOT 1ST YR MD STUDENTS 76-77	1612	
V6470	PROJTD ENROLL-TOT 1ST YR MD STUDENTS 77-78	1613	
V6480	PROJTD ENROLL-TOT 1ST YR MD STUDENTS 78-79	1614	
V6491	PROJTD ANNUAL GROWTH RATE 74-78	$(1614/1610)^{.25}-1)$	

--BY CLASS--

V6500	ENROLL-TOT 1ST YR MD STUDENTS	1382	
V6510	ENROLL-TOT MID YR MD STUDENTS	1388	
V6520	ENROLL-TOT FINAL YR MD STUDENTS	1385	

--BY SEX--

V6600	ENROLL-TOT MALE 1ST YR MD STUDENT	1380	
V6605	ENROLL-PCT FEMALE 1ST YR MD STUDENT	((1382-1380)/1380)	x100
V6610	ENROLL-TOT MALE MID YR MD STUDENT	1386	
V6615	ENROLL-PCT FEMALE MID YR MD STUDENT	((1388-1386)/1386)	x100

V6620	ENROLL-TOT MALE FINAL YR MD STUDENT	1383	
V6625	ENROLL-PCT FEMALE FINAL YR MD STUDENT	((1385-1383)/1383)	x100
V6630	ENROLL-TOT MALE MD STUDENT	1389	
V6635	ENROLL-PCT FEMALE MD STUDENT	((1391-1389)/1389)	x100

--FOREIGN MEDICAL STUDENTS--

V6700	FMS ENROLL-TOT MD STUDENTS	1394+1396+1395	
V6705	FMS ENROLL-PCT MD STUDENTS	(1394+1396+1395/1391)	x100
V6710	FMS ENROLL-TOT 1ST YR MD STUDENTS	1394	
V6715	FMS ENROLL-PCT 1ST YR MD STUDENTS	(1394/1382)	x100
V6720	FMS ENROLL-TOT MID YR MD STUDENTS	1396	
V6725	FMS ENROLL-PCT MID YR MD STUDENTS	(1396/1388)	x100
V6730	FMS ENROLL-TOT GRAD MD STUDENTS	1395	
V6735	FMS ENROLL-PCT GRAD MD STUDENTS	(1395/1385)	x100

--ETHNIC COMPOSITION--

V6800	MD STUDENTS-TOT UNDER REP MINORITY	1461	
V6805	MD STUDENTS-PCT UNDER REP MINORITY	((1435+1436/1391)	x100
V6810	MD STUDENTS-TOT CAUCASIAN MALE	1419	
V6820	MD STUDENTS-TOT CAUCASIAN FEMALE	1420	
V6830	MD STUDENTS-TOT ORIENTAL-AM MALE	1435	
V6840	MD STUDENTS-TOT ORIENTAL-AM FEMALE	1436	

--REPEATERS--

V6900	REPEATERS-PCT 1ST YR MD STUDENTS	((1490+1491)/1382)	x100
V6910	REPEATERS-TOT 1ST YR MD STUDENTS MALE	1490	
V6920	REPEATERS-TOT 1ST YR MD STUDENTS FEMALE	1491	

--WITHDRAWALS--

V7000	WITHDRL-TOT MD STUDENTS-ALL REASONS	1529	
V7005	WITHDRL-PCT MD STUDENTS-ALL REASONS	(1529/1391)	x100
V7010	WITHDRL-TOT 1ST YR-ALL REASONS	1526	
V7015	WITHDRL-PCT 1ST YR-ALL REASONS	(1526/1382)	x100
V7020	WITHDRL-TOT MID YR-ALL REASONS	1528	
V7025	WITHDRL-PCT MID YR-ALL REASONS	(1529/1371)	x100
V7030	WITHDRL-TOT FINAL YR-ALL REASONS	1527	
V7035	WITHDRL-PCT FINAL YR-ALL REASONS	(1529/1385)	x100

*** ENTERING QUALIFICATIONS ***

--GPA--

V7100	UNDERGRAD GPA-ENTERING 1ST YR MD STUDENTS	1547	
V7110	PRE MD GPA 3.6 TO 4.0-1ST YR MD STUDENTS	1530	
V7115	PRE MD GPA 3.6 TO 4.0-PCT 1ST YR MD STUDENTS	(1530/1382)	x100

V7120	PRE MD GPA 2.6 TO 3.5-1ST YR MD STUDENTS	1531	
V7125	PRE MD GPA 2.6 TO 3.5 PCT 1ST YR MD STUDENTS	(1531/1382)	x100
V7130	PRE MD GPA LESS THAN 2.6-1ST YR MD STUDENTS	1532	
V7135	PRE MD GPA LESS THAN 2.6-PCT 1ST MD STUDENTS	(1532/1382)	x100
V7140	PRE MD GPA UNKNOWN-1ST YR MD STUDENTS	1533	
V7145	PRE MD GPA UNKNOWN-PCT 1ST YR MD STUDENTS	(1533/1382)	x100

--MCAT--

V7200	MEAN MCAT SCORE SCI-1ST YR MD STUDENTS	1546	
V7210	MEAN MCAT SCORE VER-1ST YR MD STUDENTS	1543	
V7220	MEAN MCAT SCORE GEN-1ST YR MD STUDENTS	1544	
V7230	MEAN MCAT SCORE QUAN-1ST MD STUDENTS	1545	

--DEGREE STATUS--

V7300	TOT BACH-1ST YR MD STUDENTS	7537	
V7305	PCT BACH-1ST YR MD STUDENTS	(1537/1382)	x100
V7310	TOT MAS-1ST YR MD STUDENTS	1538	
V7315	PCT MAS-1ST YR MD STUDENTS	(1538/1382)	x100
V7320	TOT DOC-1ST YR MD STUDENTS	1539	
V7325	PCT DOC-1ST YR MD STUDENTS	(1539/1382)	x100
V7330	PCT ANY DEGREE-1ST YR MD STUDENTS	((1538+1539+1540)/(1537+1541))	x100
V7340	TOT OTHER DEGREE-1ST YR MD STUDENTS	1540	
V7345	PCT OTHER DEGREE-1ST YR MD STUDENTS	(1540/1382)	x100
V7350	TOT NO DEGREE-1ST YR MD STUDENTS	1541	
V7355	PCT NO DEGREE-1ST YR MD STUDENTS	(1541/1382)	x100

--UNDERGRADUATE EDUCATION--

V7400	UNDERGRAD ED-2 YRS OR LESS-1ST YR MD STUDENTS	1534	
V7405	UNDERGRAD ED-2 YRS OR LESS-PCT 1ST YR MD STUDENTS	(1534/1382)	x100
V7410	UNDERGRAD ED-3 YRS-1ST YR MD STUDENTS	1535	
V7415	UNDERGRAD ED-3 YRS-PCT 1ST YR MD STUDENTS	(1535/1382)	x100
V7420	UNDERGRAD ED-4 YRS OR MORE-1ST YR MD STUDENTS	1536	
V7425	UNDERGRAD ED-4 YRS OR MORE-PCT 1ST YR MD STUDENTS	(1536/1382)	x100

*** STUDENT AID ***

--REQUESTING--

V7500	REQ AID-TOT MD STUDENTS	1979	
V7505	REQ+RECVD AID-PCT MD STUDENTS	(1989/1979)	x100
V7510	REQ AID-TOT 1ST YR MD STUDENTS	1975	
V7515	REQ+RECVD AID-PCT 1ST YR MD STUDENTS	(1985/1975)	x100
V7520	REQ AID-TOT 2ND YR MD STUDENTS	1976	
V7525	REQ+RECVD AID-PCT 2ND YR MD STUDENTS	(1986/1976)	x100
V7530	REQ AID-TOT 3RD YR MD STUDENTS	1977	

V7535	REQ+RECVD AID-PCT 3RD YR MD STUDENTS	(1987/1977)	x100
V7540	REQ AID-TOT FINAL YR MD STUDENTS	1978	
V7545	REQ+RECVD AID-PCT FINAL YR MD STUDENTS	(1988/1978)	x100

--RECEIVING--

V7600	RECVD AID-TOT MD STUDENTS	1989	
V7610	TOT AID TO MD STUDENTS	1999	
V7615	AV AMT AID TO MD STUDENTS	(1999/1391)	x100
V7620	RECVD AID-TOT 1ST YR MD STUDENTS	1985	
V7630	TOT AID TO 1ST YR MD STUDENTS	1995	
V7635	AV AMT AID TO 1ST YR MD STUDENTS	1995/1985	
V7640	RECVD AID-TOT 2ND YR MD STUDENTS	1986	
V7650	TOT AID TO 2ND YR MD STUDENTS	1996	
V7655	AV AMT AID TO 2ND YR MD STUDENTS	1996/1986	
V7660	RECVD AID-TOT 3RD YR MD STUDENTS	1987	
V7670	TOT AID TO 3RD YR MD STUDENTS	1997	
V7675	AV AMT AID TO 3RD YR MD STUDENTS	1997/1987	
V7680	RECVD AID-TOT FINAL YR MD STUDENTS	1988	
V7690	TOT AID TO FINAL YR MD STUDENTS	1998	
V7695	AV AMT AID TO FINAL YR MD STUDENTS	1998/1988	

--NEEDING--

V7700	NEED AID-TOT MD STUDENTS	1984	
V7705	NEED+RECVD AID-PCT OF TOT MD STUDENTS	(1989/1984)	x100
V7710	NEED AID-TOT 1ST YR MD STUDENTS	1980	
V7715	NEED+RECVD AID-PCT 1ST YR MD STUDENTS	(1985/1980)	x100
V7720	NEED AID-TOT 2ND YR MD STUDENTS	1981	
V7725	NEED+RECVD AID-PCT 2ND YR MD STUDENTS	(1986/1981)	x100
V7730	NEED AID-TOT 3RD YR MD STUDENTS	1982	
V7735	NEED+RECVD AID-PCT 3RD YR MD STUDENTS	(1987/1982)	x100
V7740	NEED AID-TOT FINAL YR MD STUDENTS	1983	
V7745	NEED+RECVD AID-PCT FINAL YR MD STUDENTS	(1988/1983)	x100

--AID DISPERSED TO STUDENTS--

V7800	AID-AMT PER MD STUDENT	1999/1391	
V7810	RECVD AID-LOANS-TOT MD STUDENTS	2036	
V7815	RECVD AID-LOANS-PCT MD STUDENTS	(2036/1391)	x100
V7820	RECVD AID-SCHLSHIP-TOT MD STUDENTS	2037	
V7825	RECVD AID-SCHLSHIP-PCT MD STUDENTS	(2037/1391)	x100

*** EXPENSES ***

--TUITION, EXPENSES, & FEES--

V7900	TUIT+EXPEN PER IN ST MD STUDENT	1965
V7910	TUIT+EXPEN PER OUT ST MD STUDENT	1966
V7920	FEES+EXPEN EXCLUD TUIT PER MD STUDENT	1969
V7930	AV EXPEN PER IN ST MD STUDENT UNMARRIED	2039
V7940	AV EXPEN PER OUT ST MD STUDENT UNMARRIED	2043
V7950	TUIT+EXPEN RATIO-IN ST TO OUT ST	1965/1966

*** STUDENT SELECTION ***

--YEAR--

V8000	YR SELECTD-HS SR 73	
V8010	YR SELECTD-UNDERGRAD FR 74-75	1280
V8020	YR SELECTD-UNDERGRAD SOPH 74-75	1281
V8030	YR SELECTD-UNDERGRAD JR 74-75	1282
V8040	YR SELECTD-UNDERGRAD SR 74-75	1283
		1284

--APPLICANTS--

V8100	APPL-TOT	Division of Student Studies
V8110	APPL-TOT MALE	Division of Student Studies
V8115	APPL-PCT MALE TO TOT	Division of Student Studies
V8120	APPL-TOT FEMALE	Division of Student Studies
V8130	RATIO-MALE APPL TO ENTERING	Division of Student Studies
V8140	RATIO-FEMALE APPL TO ENTERING	Division of Student Studies
V8150	RATIO-APPL TO ENTERING	Division of Student Studies

--STANDING--

V8200	MC ACCEPT TRANS STUDENTS	1286
V8210	MC ACCEPT ADV STANDING STUDENTS	1311

*** CAREER REVIEW ***

V8300	REVIEW CAREER CHOICE AT GRADUATION	0438
V8310	REVIEW CAREER CHOICE 5 YRS AFTER GRAD 73	0439
V8330	ADVIS PROG-STUDENT RETENTION 74-75	1318
V8340	CAREER INTENT AFFECTS ADMISS DECISION	0441

APPENDIX B

19 CLUSTER SOLUTION		9 CLUSTER SOLUTION
CLUSTER #1	ARKANSAS LOUISVIL LA N ORL TENNESS MISS OKLAHOMA PR RICO MICH ST	CLUSTER #1
CLUSTER #2	TX S ANT CONN RUTGERS S DAKOTA GEORGIA S CAROL TX GALV N CAROL U VIRGIN WISCONSIN	
CLUSTER #3	MC VIRG WAYNE ST VERMONT W VIRGIN MO COLUM ALABAMA UTAH CINCIN SUNY SYR KENTUCKY NEBRASKA	CLUSTER #2
CLUSTER #4	SUNY BUF OREGON COLORADO N JERSEY ARIZONA CAL DAV FLORIDA	
CLUSTER #5	ILLINOIS SUNY DST UCLA	CLUSTER #3
CLUSTER #6	INDIANA OHIO ST U MICH WASH SEA	
CLUSTER #7	CAL IRV RUSH STONY BRK OHIO TOL LA SHREV MASS SO FLA SO ILL	CLUSTER #4
CLUSTER #8	NEW MEX MT SINAI CAL S DI	
CLUSTER #9	NEVADA E VIRGINIA	
CLUSTER #10	MAYO	CLUSTER #5

APPENDIX B

19 CLUSTER SOLUTION		9 CLUSTER SOLUTION
CLUSTER #11	U PENN CASE WST CORNELL SO CAL	
CLUSTER #12	DARTMOUTH BROWN	CLUSTER #6
CLUSTER #13	U CHICAGO J HOPKINS ROCHESTER VANDERBILT	
CLUSTER #14	MIAMI TEMPLE CHICAGO MED M C PENN CREIGHTON ST LOUIS ALBANY BOWMAN GRAY PITTSBURGH	
CLUSTER #15	GEO WASH NWESTERN HAHNEMAN HOWARD G TOWN M C WISCONSIN BOSTON JEFFERSON N Y MED LOYOLA	CLUSTER #7
CLUSTER #16	TEX TECH MINN DUL S ALABAMA	CLUSTER #8
CLUSTER #17	EINSTEIN N Y UNIV STANFORD WASH S L YALE	
CLUSTER #18	CAL S F MINN MPS	CLUSTER #9
CLUSTER #19	TX SWEST PENN ST DUKE	

Studies in Medical Education

Anderson, P. _Descriptive Study of Salaried Medical School Faculty_. December, 1975.

Johnson, D.G. and Dube, W.F. _Descriptive Study of Medical School Applicants, 1974-75_. December, 1975.

Lambdin, J.A. _Survey of How Medical Students Finance Their Education, 1974-75_. December, 1975.

Nunn, R. and Lain, L. _Classification of Medical Education Institutions_. December, 1975.

Rosenthal, J. _Medical School Programs, Resources and Financing_.

Sedlacek, W.E. _Variables Related to Increases in Medical School Class Size_. December, 1975.

Sherman, C. _Study of Medical Education: Interrelationships Between Component Variables_. December, 1975.

Additional copies of these publications may be obtained from:

Association of American Medical Colleges
Attention: Membership and Subscriptions
1 Dupont Circle N.W.
Washington, D.C. 20036